AD_____

Award Number:

W81XWH-11-1-0181

TITLE:

Mining the Immune Cell Proteome to Identify Ovarian Cancer-Specific Biomarkers

PRINCIPAL INVESTIGATOR:

Dr. Manish S. Patankar

CONTRACTING ORGANIZATION:

University of Wisconsin-Madison
Madison, WI. 53715

REPORT DATE:

November 2013

TYPE OF REPORT:

Final Report

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
November 2013	Final Report	15 February 2011- 14 October 2013

4. TITLE AND SUBTITLE
Mining the Immune Cell Proteome to Identify Ovarian Cancer-Specific Biomarkers

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-11-1-0181

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Manish S. Patankar

E-Mail: patankar@wisc.edu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
University of Wisconsin-Madison,
Madison, WI. 53715

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT In previous studies we have demonstrated that the ovarian cancer antigen CA125 specifically binds to certain sunsets of immune cell. Based on these observations we have hypothesized that proteomic and transciptomic analysis of immune cells from ovarian cancer patients will result in the identification of specific biomarkers in the immune cells. The current proposal will further investigate this hypothesis by conducting in-depth proteomic analysis of immune cells from cancer patients and healthy blood donors. Studies conducted have resulted in development of streamlined protocols for proteomic analysis of human immune cells. Proteomic analysis of human NK cells has been completed. In addition to proteomic analysis we have also compared the transcriptome of immune cells from ovarian cancer patients and healthy donors and have identified approximately 1600 genes that are differentially expressed in the patient samples. On-going research is focused on validating the proteomic and transcriptome data and demonstrating changes immune cells in response to cancer antigens.

15. SUBJECT TERMS
Ovarian cancer, immune cells, biomarker, early detection, proteomics, microarray, transcriptome

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			USAMRMC
U	U	U	UU	69	19b. TELEPHONE NUMBER *(include area code)*

Table of Contents

Introduction

Immune cells are in continuous crosstalk with tumor cells and other cells in the tumor microenvironment. This crosstalk originates from direct physical contact between the immune cells and the cells in the tumor microenvironment or through their interactions with factors (proteins, lipids and other biomolecules as well as cellular fragments such as exosomes and microparticles) released from the tumors and associated tissues. This interaction causes changes in the molecular make-up of immune cells. In the project funded by the Department of Defense (DOD) we hypothesized that these alterations in the molecular phenotype of immune cells should be explored for disease-specific biomarkers. Although initially our focus was on identifying the biomarkers by proteomics analysis, the project evolved to a stage where using modern sequencing techniques allowed us to analyze the transcriptome of immune cells to identify biomarkers that can be further validated by proteomic analysis.

Body.

The study we completed was a proof-of-principle study to demonstrate that the circulating immune cells of ovarian cancer patients be considered as repositories of cancer-specific biomarkers. To the best of our knowledge, such studies have not been conducted in ovarian cancer patients. During the first year of this study, we focused our attention on performing proteomic analysis of immune cells to establish protocols to lyse the immune cells and to conduct mass spectrometry on the lysed samples. We were successful in completing this work and were able to characterize the proteome of human natural killer cells. Our initial experiments were designed to demonstrate that the protocols we had established indeed were useful in identifying differences in immune cell proteome when the immune cells were subjected to stimulation. We used a minimalist approach and conducted proteomic analysis of NK cells isolated from the blood of three healthy donors. The NK cells were stimulated in vitro with IL-2. Using high resolution mass spectrometry, we were able to show several significant changes in the proteome of NK cells that were stimulated with IL-2 as compared to naïve unstimulated freshly isolated NK cells.

These initial experiments not only allowed us to develop the protocols for proteomic analysis but also provided support to our hypothesis that soluble factors such as cytokines can indeed cause significant alterations in the proteomes of immune cells. While it is predictable that IL-2 would cause changes in NK cells via its signaling through the IL-2 receptor pathway, we found evidence of proteomic changes in pathways and genes that are not normally found to be associated or affected by IL-2 receptor stimulation (abstracts and paper published on this topic is included in **Appendix 1-4**, respectively).

Based on these results we predict that under similar situation when the immune cells are under the influence of cytokines and other modulating factors originating from ovarian cancer cells or from the ovarian tumor microenvironment, major proteomic changes in the immune cells should be expected. The challenge now is to exploit the knowledge and

conclusions from this proof-of-concept study to identify biomarkers that can be used for screening as well as monitoring of ovarian cancer.

We will revert to this point and discuss other advances we have made in identifying immune cell-based biomarkers of ovarian cancer in a subsequent paragraph. But first we will also discuss how the data we collected on the proteomic analysis of IL-2 stimulated NK cells will be helping us with understanding mechanism of an immunotherapeutic agent that can target human ovarian tumors.

In a separate project we are investigating huKS-IL2, an immunocytokine (antibody-IL2 chimera) that targets Epithelial Cell Adhesion Molecule (EpCAM) that is overexpressed by ovarian and other tumors. huKS-IL2 recruits NK and other innate immune cells and facilitates cytolysis of ovarian cancer cells. In recent work we showed that the IL-2 molecules of huKS-IL2 engage the IL-2 receptor on NK cells and facilitate the formation of an immunological synapse- a step that is essential for the NK cells to lyse tumor targets. The immunocytokines are engineered to have two molecules of IL-2 per molecule of antibody. Thus, each molecule of huKS-IL2 has two IL-2 molecules. Based on our preliminary data, it is clear that the two IL-2 molecules crosslink IL-2 receptors on NK cells. With the immunocytokines we have a situation where the innate immune cells are stimulated not only via the IL-2 receptors but also through engagement of Fc receptors that are engaged by the antibody portion of huKS-IL2.

The protocols we have developed for proteomic analysis of NK cells will be used in our studies with huKS-IL2 to determine changes in proteome of the immune cells as a result of activation of two separate receptor systems- the IL-2 and the Fc receptors. These studies will provide important data on the molecular mechanisms of huKS-IL2 and other similar immunocytokines.

Now we will revert back to our further studies on identifying ovarian cancer-specific biomarkers in circulating immune cells. While the proteomic analysis provided useful information, we realized that the procedures required for this analysis are generally complex and require significant numbers of immune cells. We therefore considered the possibility of gaining biomarker information by conducting a transcriptomic analysis of immune cells. Such analysis will provide us with direct evidence of genes that are differentially regulated in immune cells from cancer patients. We reasoned that once the differentially expressed genes are identified in the immune cells, proteomic (ELISA, flow cytometry, etc.) could be used to validate the results at the protein level and to develop screening or diagnostic tests for ovarian cancer.

With this intention, we conducted a study where RNA samples were obtained from immune cells isolated from five healthy women and five ovarian cancer patients with stage III or IV ovarian cancer. This approach was very successful as it resulted in identification of several hundred genes that were differentially expressed in the two cohorts. From this data we have identified a panel of five genes that were most differentially expressed. We are currently continuing our work validating these molecular signatures in immune cells isolated from age matched healthy donors, patients with

benign gynecologic disease and advanced ovarian cancer patients (twenty subjects in each group).

With the microarray analysis showing positive results, we further explored if the use of modern RNA sequencing techniques can be used to further enhance the chances of identifying cancer-specific biomarkers in the immune cell transcriptome. RNAseq is a powerful technique for high-throughput sequencing of nucleic acids and has significant advantages over conventional microarray experiments. The information obtained from RNAseq analysis not only provides a quantitative assessment of gene expression but also provides information about potential mutations, expression of splicing variants and expression of microRNAs.

Considering the benefits, we performed RNAseq analysis of RNA samples isolated from three healthy donors and three patients with advanced stage ovarian cancer. The sequence data obtained through these experiments is enormous and we are mining this data to identify transcriptome signatures unique to the immune cells from ovarian cancer patients. Our initial bioinformatics analysis conducted in collaboration with Dr. Jesus Gonzalez-Bosquet has led us to identify approximately 1300 unique genes that are differentially expressed in cancer patient's immune cells. Experiments are continuing to validate specific genes from the RNAseq experiments in the healthy, benign gynecologic and cancer patients referenced above. We are also comparing the microarray dataset with the dataset from the RNAseq experiments to validate the data.

In continuing experiments we are using a mouse model to further demonstrate the power of using the immune cell transcriptome as repository for cancer-specific biomarkers. Mouse ovarian tumors are implanted in C57/BL6 mice and immune cells are being drawn at regular intervals from these animals. Transcriptomes of the immune cells will be characterized by RNAseq and changes corresponding different stages of tumor growth will be determined. This experimental model will further support our overall concept and also provide us with an experimental paradigm to conduct advanced well-controlled experiments.

Finally, we would also wish to report that our hypothesis of finding biomarkers in immune cells can also be extended to other diseases. For example, we have reported our work on proteomic changes in immune cells derived from women with norml pregnancies and those with preeclampsia. Conceptual understanding for this work came from the project funded by the DOD and we were able to use some of the control samples for this preeclampsia study that we had already obtained for the studies proposed in our DOD proposal. A paper on this data on immune cells from preeclampsia patients is published and is attached as **Appendix 5**.

Key Accomplishments:

1. Proteomic profiling of human NK cells and identification of novel signaling pathways and molecules that are triggered in these immune cells following activation by IL-2.

2. Identification of a panel of potential biomarkers by the microarray analysis of the transcriptome of immune cells from ovarian cancer patients and healthy donors.
3. Completion of the RNAseq analysis of the transcriptome of immune cells from ovarian cancer patients and healthy donors.
4. Identification of several gene signatures through RNAseq and generation of a large database from this data that can be queried and mined for future studies.
5. Validation experiments are underway on immune cells from healthy donors, ovarian cancer patients and women with benign gynecologic pathology.
6. The success of our proof-of-concept studies has also allowed us to test if biomarker signatures in immune cells isolated from patients benign pathologies. We have conducted RNAseq analysis on RNA samples isolated from circulating immune cells from normal pregnancies and those with preeclampsia. Thus the funding provided to us through this DOD grant mechanism will not only advance research in ovarian cancer but also other cancers and benign conditions.

Reportable outcomes:

1. Publication of a manuscript in the proteomic analysis of NK cells in the Journal of Proteomics.
2. A R01 proposal is currently under construction and is planned to be submitted to NIH by the February 5, 2014 deadline.
3. We obtained additional funding from the Department of Obstetrics and Gynecology and the University of Wisconsin-Madison Paul P. Carbone Cancer center to support this project.
4. Drs. Shitanshu Uppal and Erin Medlin are fellows in the PI's division who are undergoing training to become Gynecologic Oncologists. As part of their fellowship requirements, Drs. Uppal and Medlin conducted one year of laboratory research. During his research year, Dr. Uppal conducted the transcriptome analysis using the microarrays. Dr. Medlin is currently working under the PI's mentorship to identify changes in immune cell transcriptome in response to ovarian tumors in the mouse model. Thus the funding from DOD has allowed us to train two future gynecologic oncologists.
5. Two graduate students were funded on this project. One student (Dr. Di Ma) received her PhD partly based on the proteomic analysis of NK cells (**Appendix 1-4**)

Conclusions.

The proteomic, transcriptome analysis collectively provide strong evidence supporting the presence of biomarkers for ovarian cancer in circulating immune cells. Our ongoing studies will be focused on consolidating and verifying these initial promising results with the intent of developing novel diagnostic assays for ovarian cancer.

Bibliography

Abstracts

1. Ma, D., Kapur, A., Felder, M., Patankar, M.S., Li, L. (2011) Characterization and comparative analysis of proteomic profiles of leukemic and primary human NK cells. Presented at the 2011 annual meeting of the American Society for Mass Spectrometry

2. Ma, D., Cao, W., Kapur, A., Scarlett, C.O., Patankar, M.S., Li, L. (2012) Comparative analysis of global proteome of naïve and IL-2 stimulated human NK cells. Presented at the 2012 annual meeting of the American Society for Mass Spectrometry

Publications

1. Ma D, Cao W, Kapur A, Felder M, Scarlett CO, Patankar MS, Li L: Differential expression of proteins in naive and IL-2 stimulated primary human NK cells identified by global proteomic analysis. *J Proteomics* 2013.
2. Tyler C, Kapur, A., Felder, M., Belisle, JA, Trautman, C, Gubbels JAA, Connor, JP, Patankar, MS: The mucin MUC16 (CA125) binds to NK cells and monocytes from peripheral blood of women with healthy pregnancy and preeclampsia. *American Journal of Reproductive Immunology* 2012, 68:28-37.
3. Cao W, Ma D, Kapur A, Patankar MS, Ma Y, Li L: RT-SVR+q: A strategy for post-mascot analysis using retention time and q value metric to improve peptide and protein identifications. *J Proteomics* 2011, 75:480-490.

Appendix 1

Abstract of work presented at the 2011 annual meeting of the American Society for Mass Spectrometry

Characterization and Comparative Analysis of Proteomic Profiles of Leukemic and Primary Human NK Cells

Di Ma[1], Arvinder Kapur[2], Mildred Felder[2], Manish S Patankar[2] and Lingjun Li[1]

[1]*School of Pharmacy, University of Wisconsin-Madison, 777 Highland Ave., Madison, WI 53705*

[2]*Department of Obstetrics and Gynecology, University of Wisconsin-Madison, 600 Highland Ave, Madison, WI 53792*

INTRODUCTION

As part of the innate immune system, natural killer (NK) cells detect and lyse tumor and virus infected cells. NK leukemia cell line NKL was initially established from patients with NK cell lymphoma. The morphology of NKL cells resembles that of primary NK cells, and they are used as "model cell lines" in NK cell research. NKL cells and primary NK cells, however, differ substantially in the expression of cell surface receptors and intracellular signal transduction proteins. To obtain useful knowledge on the functional relevance of these two cell types, we characterize and compared the proteome of NKL cells and primary NK cells isolated from healthy donors by mass spectrometry-based quantitative proteomics.

METHODS

NKL cells and primary NK cells were harvested and subjected to subcellular fractionation by differential centrifugation followed by protein extraction. The protein concentration was determined by BCA assay. The detergent in protein sample was removed by acetone precipitation and the proteins were redissolved in 8M urea. To reduce sample complexity, multidimensional liquid chromatography was employed prior to MS analysis. The tryptic digests of NKL cell lysates were separated in the first dimension off-line using high-pH RPLC and fractions were collected and subjected to the second dimension nanoflow RPLC on-lined coupled to MS/MS analysis. The resulting spectra were searched against Mascot for protein identification. The comparative analysis of NKL cell proteome and primary NK cell proteome is conducted via spectral counting.

PRELIMINARY DATA

Two most commonly used cell lysis buffers, RIPA buffer and NP-40 buffer were used for protein extraction in NKL cells and the efficiency of each buffer was compared based on protein concentration measured by BCA assay. It was shown that RIPA buffer gives a much higher protein yield (8.1µg/µl) compared to NP-40 buffer (3.01µg/µl). To remove the detergent in protein sample, acetone precipitation was conducted. However, it was

difficult to redissolve the protein pellet after acetone precipitation. To solublize proteins, either 8M urea or ProteaseMAX™ surfactant was added to assist trypsin digestion, and protein identification results showed that the two methods were comparable in improving protein solubility and enhancing in-solution tryptic digestion. To reduce the sample complexity, we employed 2D RP-RPLC prior to tandem MS experiments. To optimize the condition in first dimension high-pH RPLC, different sample loading amount (20µg *vs*. 50µg) and different fraction collection intervals (3min *vs.* 2min) were compared. It was determined that injecting 50µg of protein digests onto first dimension high-pH RPLC resulted in 25% increase in the number of protein identifications. A comparison of 2 minute versus 3 minute fraction collection interval revealed that 2 minute/fraction was more favorable for downstream LC-MS/MS and the number of proteins identified was increased by 30%. Using the optimized condition for protein extraction and sample separation, a total of 1054 proteins were identified from four replicates of NKL cell lysates, and the sensitivity of our method was demonstrated by the detection of a low abundance protein HLA-DR alpha chain which is only expressed in NKL cells but not in primary NK cells. This and other differences in the proteomes highlight the potential variations in the cytotoxic and cytokine producing functions of the primary NK and the leukemic NKL cells. The proteome of primary NK cells isolated from healthy donors will be investigated using the same protocol, and the comparative analysis of NKL cell proteome and primary NK cell proteome will be performed via label-free spectral counting methodology.

NOVEL ASPECT

The combined use of subcellular fractionation and 2D RP-RPLC to reduce sample complexity prior to MS analysis of immune cell proteomes allows us to exhaustively study the alterations in the NK cell proteome in response to tumor antigens, and chemo and immunotherapeutic regimens.

Appendix 2: Poster presented at the 2012 Annual meeting of the American Society for Mass Spectrometry

Comparative Analysis of the Global Proteome of Naïve and IL-2 Stimulated Human NK Cells

Di Ma[1], Weifeng Cao[2], Arvinder Kapur[3], Cameron O Scarlett[1] Manish S Patankar[3] and Lingjun Li[1, 2]

[1]School of Pharmacy, University of Wisconsin-Madison, 777 Highland Ave., Madison, WI 53705
[2]Department of Chemistry, University of Wisconsin-Madison, 777 Highland Ave., Madison, WI 53705
[3]Department of Obstetrics and Gynecology, University of Wisconsin-Madison, 600 Highland Ave, Madison, WI 53792

THE UNIVERSITY of WISCONSIN MADISON

Overview

Purpose:

- To develop a mass spectrometry workflow for protein characterization and quantitative analysis of primary NK cells .
- To compare the proteome of naïve and IL-2 stimulated human NK cells.

Methods:

- Protein was extracted with RIPA buffer with brief sonication.
- Prior to MS analysis, 2D RP-RPLC was employed to reduce sample complexity .
- Quantitative proteomic analyses of naïve and IL-2 activated NK cells were performed by spectral counting.

Results:

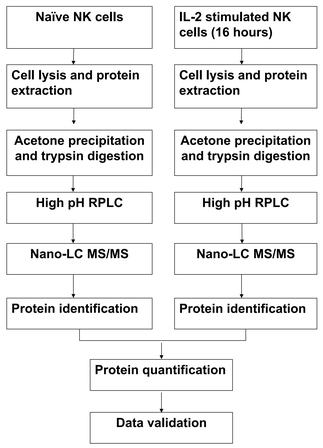
Proteomic profile of naïve and IL-2 activated human NK cells were investigated and compared . Quantitative analysis revealed a list of more than 400 proteins with more than 2-fold up or down-regulation upon IL-2 stimulation.

Introduction

Natural Killer (NK) cells efficiently cytolyse tumors and virally infected cells. The development and functional activity of NK cells are regulated by many cytokines, including IL-2 which stimulates the proliferation of NK cells and increases NK cells activity. Although how IL-2 mediates its effects has been investigated, little is known about the alterations in the global NK cell proteome following IL-2 activation. With the development of NK cell-based immunotherapies relying on activation via IL-2, it is important to conduct an exhaustive proteomic analysis of the NK cell to delineate molecular pathways that may impinge or accentuate the immune responses. We therefore characterized the proteome of naïve and IL-2 activated primary NK cells by mass spectrometry-based quantitative proteomics.

Methods

Workflow



Primary NK Cell Isolation

Primary NK cell were isolated from healthy donors using RosetteSep® Human NK Cell Enrichment Cocktail.

Sample Preparation

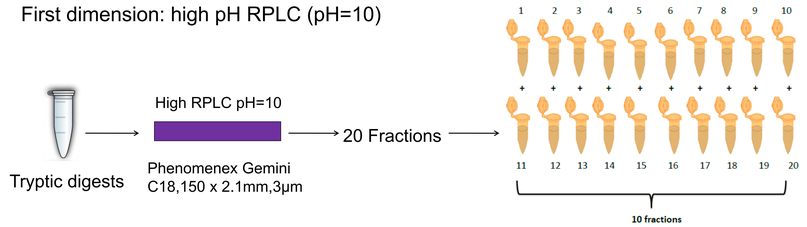
- Proteins were extracted from naïve or IL-2 treated primary NK cells using RIPA lysis buffer.
- Protein samples were cleaned by acetone precipitation.
- Proteins were digested by trypsin.

Sample Separation

Tryptic digests were separated by high pH reversed phase liquid chromatography offline prior to nanoLC-MS/MS analysis.

Multidimensional Liquid Chromatography (MDLC)

First dimension: high pH RPLC (pH=10)



The scheme of employing high RPLC for orthogonal separation prior to MS analysis.

Instrumentation

LC-MS/MS

amaZon ion trap mass spectrometer equipped with Waters nanoAcquity.
NanoLC separation: Waters 1.7µm BEH130 100µm x 100mm analytical column.

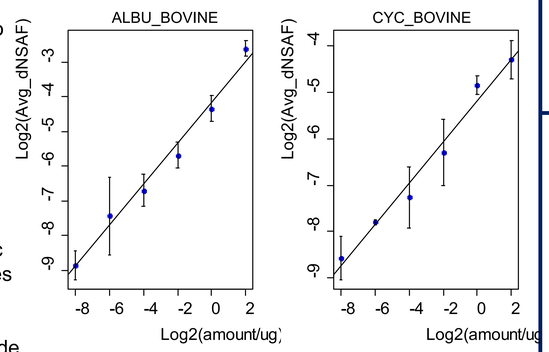
Data Analysis

Protein identifications were performed by Mascot database searching. The embedded Mascot percolator was utilized to improve sensitivity of identifications. Each biological replicate has three technical replicates, and proteins detected in at least two technical replicates were selected for protein identification. Proteins identified in all three biological replicates were selected for quantification by spectral counting.

Results

Validation of Protein Quantification method

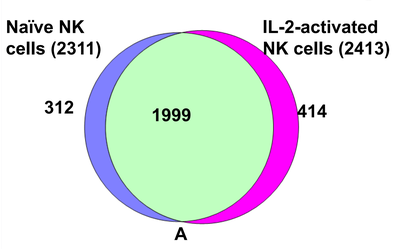
Mascot percolator was applied to facilitate and validate protein identification. To test if the application of percolator would affect the accuracy of protein quantification. The protein quantification method was validated by spiking 4µg, 1µg, 0.25 µg, 62.5ng, 15.625ng, 3.91ng of BSA and cytochrome c into 10µg digests from cell lysates of leukemic natural killer cells. The dynamic range of protein standards is 3 orders of magnitude. dNSAF was calculated as previously described by Zhang et al. [1].



1. Zhang Y et al., Anal Chem. 82, 2272-81 (2010)

Figure 1. Linear regression between dNSAF values and known protein amounts. Log2-transformed of dNSAF values were plotted as a function of log2-transformed proteins amounts in micrograms.

Protein Identification and Quantification



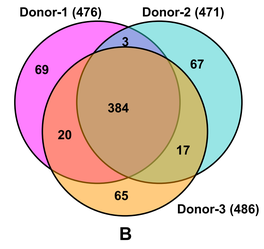


Figure 2. Summary of protein identification and quantification. (A) Number of proteins identified from all 3 donors. (B) Number of up or down-regulated proteins in response to IL-2 in different donors.

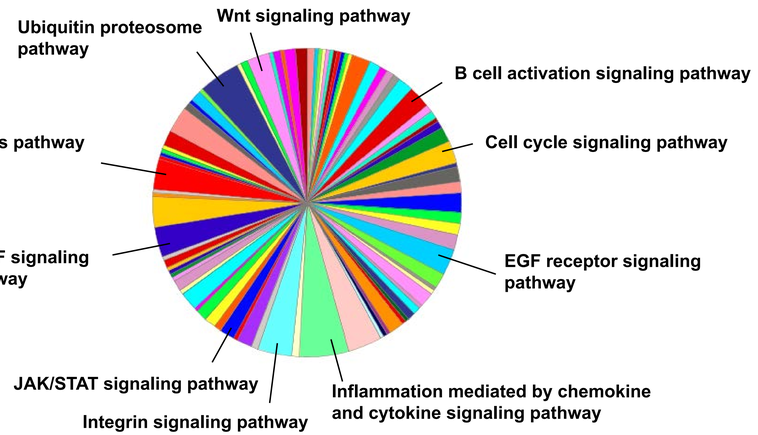


Figure 3. PANTHER pathway analysis. The functions of 384 differentially expressed proteins in all three donors were searched in PANTHER database. PANTHER pathway analysis revealed 89 pathways from a total of 254 protein hits.

Conclusions and Future Directions

In summary, we have developed a method for the proteomic analysis in human primary NK cells. We successfully employed 2D LC to reduce the sample complexity prior to MS analysis. To improve protein identification, Mascot percolator was employed and 2311 and 2413 proteins could be identified from naïve or IL-2-activated NK cells respectively. Quantitative analysis revealed a list of more than 400 proteins that are up or down-regulated in IL-2 signaling, which will be further examined and validated in future work.

Acknowledgements

- UW-Madison School of Pharmacy Analytical Instrumentation Center
- Department of Defense Pilot Award (W81XWH-11-1-0181)

JOURNAL OF PROTEOMICS 91 (2013) 151–163

Available online at www.sciencedirect.com

ScienceDirect

www.elsevier.com/locate/jprot

Differential expression of proteins in naïve and IL-2 stimulated primary human NK cells identified by global proteomic analysis

Di Ma[a], Weifeng Cao[b,1], Arvinder Kapur[c,1], Mildred Felder[c], Cameron O. Scarlett[a], Manish S. Patankar[c,*], Lingjun Li[a,b,**]

[a]School of Pharmacy, University of Wisconsin-Madison, 777 Highland Ave., Madison, WI 53705, USA
[b]Department of Chemistry, University of Wisconsin-Madison, 1101 University Ave., Madison, WI 53706, USA
[c]Department of Obstetrics and Gynecology, University of Wisconsin-Madison, 600 Highland Avenue, Madison, WI 53792, USA

ARTICLE INFO

Article history:
Received 24 March 2013
Accepted 17 June 2013
Available online 25 June 2013

Keywords:
NK cells
IL-2 signaling
Mass spectrometry pathways
Proteomics
Lab-free quantification
Multi-dimensional separation

ABSTRACT

Natural killer (NK) cells efficiently cytolyse tumors and virally infected cells. Despite the important role that interleukin (IL)-2 plays in stimulating the proliferation of NK cells and increasing NK cell activity, little is known about the alterations in the global NK cell proteome following IL-2 activation. To compare the proteomes of naïve and IL-2-activated primary NK cells and identify key cellular pathways involved in IL-2 signaling, we isolated proteins from naïve and IL-2-activated NK cells from healthy donors, the proteins were trypsinized and the resulting peptides were analyzed by 2D LC ESI-MS/MS followed by label-free quantification. In total, more than 2000 proteins were identified from naïve and IL-2-activated NK cells where 383 proteins were found to be differentially expressed following IL-2 activation. Functional annotation of IL-2 regulated proteins revealed potential targets for future investigation of IL-2 signaling in human primary NK cells. A pathway analysis was performed and revealed several pathways that were not previously known to be involved in IL-2 response, including ubiquitin proteasome pathway, integrin signaling pathway, platelet derived growth factor (PDGF) signaling pathway, epidermal growth factor receptor (EGFR) signaling pathway and Wnt signaling pathway.

Biological significance
The development and functional activity of natural killer (NK) cells is regulated by interleukin (IL)-2 which stimulates the proliferation of NK cells and increases NK cell activity. With the development of IL-2-based immunotherapeutic strategies that rely on the IL-2-mediated activation of NK cells to target human cancers, it is important to understand the global molecular events triggered by IL-2 in human NK cells. The differentially expressed proteins in human primary NK cells following IL-2 activation identified in this study confirmed the activation of JAK–STAT signaling pathway and cell proliferation by IL-2 as expected, but also led to the discovery and identification of other factors that are potentially important in IL-2 signaling. These new factors warrant further investigation on their potential roles in modulating NK cell biology. The results from this study suggest that the activation of NK cells by IL-2 is a dynamic process through which proteins with various functions are regulated.

* *Corresponding author.*
** *Correspondence to:* L. Li, School of Pharmacy, University of Wisconsin-Madison, 777 Highland Avenue, Madison, Wisconsin 53705-2222, USA. Tel.: +1 608 265 8491; fax: +1 608 262 5345.
E-mail addresses: patankar@wisc.edu (M.S. Patankar), lli@pharmacy.wisc.edu (L. Li).
[1] These authors contributed equally to the manuscript.

1874-3919/$ – see front matter
http://dx.doi.org/10.1016/j.jprot.2013.06.024

Such findings will be important for the elucidation of molecular pathways involved in IL-2 signaling in NK cells and provide new targets for future studies in NK cell biology.

1. Introduction

Natural killer (NK) cells are large granular lymphocytes generated in bone marrow that make up 5–15% of the peripheral blood lymphocytes (PBLs) [1]. NK cells mediate important innate immune responses that protect against viral infections and cancer. NK cells directly kill target cells by recruiting granzyme and perforin containing secretory granules to the immunological synapse with target cells and also release proinflammatory cytokines such as interferon gamma (INF-γ) and tumor necrosis factor (TNF). The immunologic activities of NK cells are controlled by intracellular signal transduction mediated by specific cytokines and a complex crosstalk between activating and inhibitory receptors [2].

IL-2, one of the first cytokines discovered, is a 15.5 kDa protein that stimulates the proliferation of T cells and NK cells [3]. In NK cells, IL-2 also augments cytotoxic function [4,5] (conventionally referred to as the Lymphokine-Activated Killer (LAK) cell activity) and induces IFN-γ production. IL-2 mediates its effects through interaction with a cell surface receptor complex consisting of IL-2Rα (CD25) and IL-2Rβ (CD122) and IL-2Rγ_c chain (CD132) [6]. Upon activation, IL-2Rβ and γ_c chain phosphorylate two Janus tyrosine-kinases (JAKs), JAK1 and JAK3, which are required for IL-2 signaling in T cells and NK cells [7–10]. Phosphorylation of JAK1 and JAK3 leads to recruitment and activation of Signal Transducers and Activators (STATs), a family of transcription factors that contribute to the diversity of cytokine responses. Following activation, STAT1, STAT3, and STAT5 translocate to the nucleus and activate target genes [11]. In addition to the JAK–STAT pathway, IL-2 also mediates signal transduction via protein kinase C (PKC), mitogen-activated protein kinase (MAPK)/extracellular signal-regulated protein kinase (ERK) and NF-κB [12]. Rapid activation of MAPK kinase (MKK)/ERK pathway by IL-2, for example, contributes to the generation of LAK activity, IFN-γ expression, and increased surface expression of CD26 and CD69 on NK cells [13]. Additionally, IL-2-mediated activation of NF-κB is responsible for up-regulation of perforin [14]. Recent studies have also shown that the non-apoptotic functions of caspases also contribute to the IL-2 mediated proliferation of NK cells [15].

All of these studies suggest that IL-2 initiates a multifactorial signaling response that collectively regulates the biological activity of NK cells. In-depth studies that utilize appropriate "omic" approaches will provide important strategies to identify the global molecular events triggered by IL-2 in human NK cells. Given the increasing use of immunotherapeutic strategies that rely on the IL-2-mediated activation of NK cells to target human cancers the impact of cutting-edge approaches to fully characterize IL-2 signaling could be especially relevant. Some IL-2 centered strategies have already been approved by the FDA for the treatment of metastatic melanoma [16] and renal cell carcinoma [17]. Intense research (basic, translational, and clinical) is also underway on the use of IL-2 and IL-2-antibody conjugates (immunocytokines) to boost the anti-cancer activities of NK [18–21]. Our initial studies on the characterization of the global proteome of naïve and IL-2-stimulated human NK cells reveal a large number of proteins exhibiting changes in expression levels upon IL-2 stimulation. Previous studies have conducted gel-based analysis of high abundance proteins from the membrane and secretory lysosome of NK cells [22–26]. The current study is the first to utilize a non-gel-based LC–MS/MS shotgun approach to carefully study the IL-2-induced changes in the global proteome of NK cells. More than two thousand proteins were identified from cell lysates of NK cells, representing the largest protein catalog reported to date. Label-free quantification by spectral counting was used to identify 383 NK cell proteins that were significantly up- or down-regulated in response to IL-2 stimulation.

2. Experimental methods

2.1. Materials

Protein standards, bovine serum albumin (BSA), bovine cytochrome c (bCYC) and equine myoglobin (eMYG) were purchased from Sigma-Aldrich (St. Louis, MO). RIPA buffer was purchased from Pierce (Rockford, IL). Urea and ammonium bicarbonate were from Fisher Scientific (Fair Lawn, NJ, USA). Ammonium formate and iodoacetamide (IAM) were purchased from Sigma-Aldrich (St. Louis, MO). Dithiothreitol (DTT) and sequencing grade modified trypsin was purchased from Promega (Madison, WI). High quality LC–MS grade and Optima grade solvents (ACN and water) were purchased from Fisher Scientific (Fair Lawn, NJ, USA). Recombinant human IL-2 was purchased from Peprotech (Rocky Hill, NJ).

2.2. Sample collection

Human primary NK cells were isolated from three healthy donors. Informed consent was obtained from all three blood donors recruited and the study was approved by the Institutional Review Board at the University of Wisconsin-Madison. NK cells from the blood samples were isolated by negative selection. The RosetteSep NK cell isolation kit (Stem Cell Technologies) was used and NK cell purification was conducted according to the manufacturer's protocol and the purity of the isolated NK cells was determined by monitoring of CD3, CD16, CD56, and NKp46 expression via flow cytometry as described in our previous work [27–29]. The purified NK cells from each donor were divided equally into two groups. Each group was cultured in medium supplemented with or without IL-2 for 16 h. Following IL-2 (300 U/ml) stimulation, the activation status of NK cells was confirmed by monitoring elevation in CD69 expression (Fig. 1). NK cells were washed three times with ice-cold PBS, resuspended in 200 μL RIPA, sonicated for 20 s and incubated on ice for 20 min. Cellular debris was removed by centrifugation for 30 min at 16,100 ×g at 4 °C. Supernatants were collected and protein concentrations were measured using a BCA protein assay kit (Pierce). Acetone (chilled to −80 °C) was added gradually

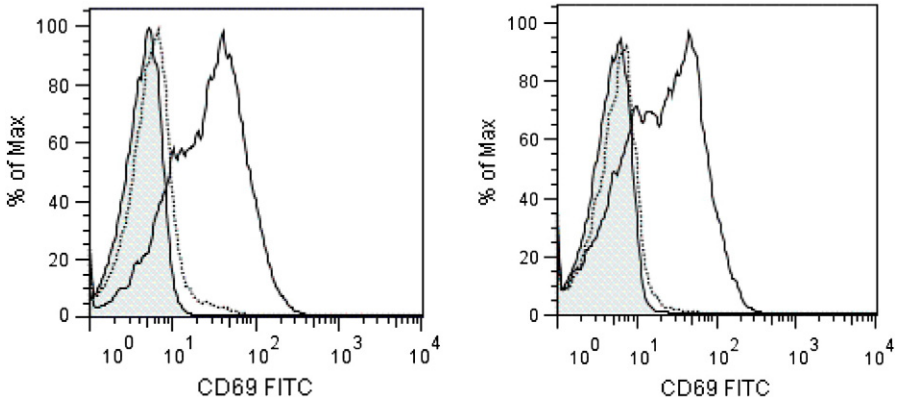


Fig. 1 – NK cell activation. NK cells isolated from the peripheral blood of healthy donors were stimulated for 16 h with IL-2. Activation status of the NK cells was determined by monitoring increase in the CD69 levels on the surface of the NK cells using flow cytometry. Data shown is for NK cells isolated from two different healthy donors. Control NK cells (dotted line) and IL-2 stimulated cells (solid line) were labeled with FITC-conjugated anti-CD69 antibody. Shaded area depicts binding of non-specific murine IgG antibody.

(with intermittent vortexing) to the protein extract to a final concentration of 80% (v/v). The solution was incubated at −20 °C for 60 min and centrifuged at 16,100 ×*g* for 15 min. The supernatant was decanted, and the pellet was carefully washed twice using cold acetone to ensure efficient removal of detergent. Residual acetone was evaporated at ambient temperature.

2.3. *Proteolysis*

All protein samples were denatured with 8 M urea in 25 mM ammonium bicarbonate buffer, and reduced by incubating with 50 mM DTT at 37 °C for 1 h. The reduced proteins were alkylated for 1 h in darkness with 100 mM iodoacetamide. The alkylation reaction was quenched by adding DTT to a final concentration of 50 mM. The samples were diluted to a final concentration of 1 M urea. Trypsin was added to the sample at a 30:1 protein to trypsin mass ratio. The sample was incubated at 37 °C overnight.

2.4. *Off-line first dimension high pH RPLC*

Tryptic digests (38 μg) from each sample were injected onto a Waters Alliance HPLC (Waters Corp., Milford, MA) with a high pH-stable RP column (Phenomenex Gemini C18, 150 mm × 2.1 mm, 3 μm) at a flow rate of 150 μL/min. The peptides were eluted with a gradient from 5 to 45% solvent B over 45 min (solvent A: 100 mM ammonium formate, pH 10; solvent B: acetonitrile (ACN)). Fractions were collected every 2 min. Twenty fractions were collected from the first dimensional RPLC at pH 10, and then every two fractions with equal collection time interval were pooled, one from the early eluted section and the other from the later eluted section as previously described [30]. The ten pooled fractions were dried by Speedvac and reconstituted in 30 μL of 0.1% formic acid. 5 μL of each fraction was subjected to nanoLC–MS/MS.

2.5. *LC–ESI ion trap mass spectrometry and MS/MS analysis*

Ten pooled fractions collected from high pH RPLC were analyzed using amaZon ETD ion trap mass spectrometer (Bruker Billerica, MA) equipped with Waters nanoAcquity UPLC (Waters Corp., Milford, MA). For the chromatographic separation, solvent A consisted of 0.1% formic acid in water and solvent B consisted of 0.1% formic acid in ACN. 5 μL of each sample was injected onto a Waters Symmetry C18 5 μm 180 μm × 20 mm precolumn at a flow rate of 5 μL/min for 5 min at 95% A/5% B, followed by peptide separation performed on Waters BEH130 1.7 μm C18 100 μm × 100 mm analytical column using gradient from 0 to 45% solvent B at 300 nL/min over 90 min. Acquisition of precursor ions and MS/MS spectra was performed using the parameters as indicated below: Smart parameter setting (SPS) was set to 700 *m/z* and compound stability and trap drive level were set at 100%. Dry gas temperature, 125 °C; dry gas, 4.0 L/min; capillary voltage, −1300 V; end plate offset, −500 V; MS/MS fragmentation amplitude, 1.0 V; and Smart fragmentation set at 30–300%. Data were generated in data-dependent mode with strict active exclusion set after two spectra and released after 1 min. MS/MS spectra were obtained via collision induced dissociation (CID) fragmentation for the six most abundant MS ions. For MS generation the ion charge control (ICC) target was set to 200,000; maximum accumulation time, 50.00 ms; one spectrometric average; rolling average, 2; acquisition range of 300–1500 *m/z*; and scan speed (enhanced resolution) of 8100 *m/z* s^{-1}. For MS/MS generation the ICC target was set to 300,000; maximum accumulation time, 50.00 ms; two spectrometric averages; acquisition range of 100–2000 *m/z*; and scan speed (Ultrascan) of 32,000 *m/z* per second.

2.6. *Database search*

MS/MS spectra were converted into Mascot Generic Format (.mgf) files by DataAnalysis (Ver 4.0, Bruker Daltonics Billerica, MA). Deviations in parameters from the default Protein Analysis in DataAnalysis were as follows: intensity threshold, 1000; maximum number of compounds, 1E9; and retention time window 0.001 min. The resulting mgf files were then searched against a home-built Human SwissProt database (SwissProt_2011_12.fasta, 533,657 entries plus 3 standard proteins, BSA, bCYC and eMYG) with Mascot 2.3.02. The searching parameters and criteria were

set as the following: tryptic digestion, maximum 2 missed cleavages, carbamidomethylation of cysteine as the fixed modification, oxidation of methionine as the variable modification, peptide mass tolerance of 1.2 Da, fragment mass tolerance of 0.6 Da, 2+, 3+ and 4+ chosen for charge state. In this study a simultaneous target-decoy search strategy (automatic decoy search) was adopted to facilitate false discovery rate (FDR) estimation. With simultaneous target-decoy search, a MS/MS spectrum is simultaneously searched against a protein sequence from the target database and its scramble version in decoy database. Therefore, FDR can be calculated from decoy hits and target hits. We applied Mascot Percolator to improve peptide and protein identification. Mascot Percolator has been well developed [31] and embedded into Mascot search engine. Mascot Percolator is a well performing machine learning method which constructs a support vector machine by using Mascot search parameters and results to re-rank peptide or protein identification. With Mascot Percolator, some low quality MS/MS spectra were re-searched to produce reliable peptide identification. We check "Percolator" option on Mascot results and control the resulting FDR at ~1%. Set "a bold red peptide required" for protein assembly. Since there were three technical replicates for each sample, only proteins identified in at least two out of three replicates were considered.

2.7. Protein quantification

Given the biological variation and technical variation across datasets, only a subset of identified proteins is qualified for quantification. It is important to select the appropriate amount of quantifiable proteins. Since we prepared three biological samples and each sample has three replicates, we set the following criteria to select quantifiable proteins: 1) a protein is quantifiable if it can be detected in at least two of three technical replicates; and 2) it can be detected in all three biological replicates. An additional complication results from protein sequence conservation. Thus in our analysis the spectral counts observed in the mass spectrometer included the tryptic peptides shared between proteins. To address this limitation we calculate the distributive normalized spectral abundance factor (dNSAF) for each quantifiable protein within a chromatographic run using the following formula [32]:

$$\mathrm{dNSAF}_i = \frac{\mathrm{dSAF}_i}{\sum_{i=1}^{N} \mathrm{dSAF}_i} \tag{1}$$

$$\mathrm{dSAF}_i = \frac{\mu SpC_i + \frac{\mu SpC_i}{\mathrm{M}} \times sSpC_i}{\sum_{m=1} \mu SpC_m / \mathrm{M}_i} \tag{2}$$

where dSAF is the distributive spectral abundance factor for a given protein, μSpC is the spectral counts of unique peptides associated with this protein while $sSpC$ is the spectral counts of shared peptides associated with this protein. M is the mono-isotopic mass of protein. The definition of "unique" peptide is the peptide whose sequence matched only one protein whereas "shared" peptide means the peptide whose sequence is shared by multiple proteins. Another limitation of our approach is the variability in peptide detection between runs. In some cases a peptide may not be observed or identified in some runs resulting in a zero count value for that peptide in a particular run. The following method was used to impute spectral counts with zero value. First, if a protein had only one zero spectral count out of three technical replicates, we calculated the average value of spectral counts of this protein in all three replicates, and then replaced the zero spectral count with the average value. Next, in the case of zero spectral counts in all three replicates, we followed the method as previously described [33] to determine a fraction value within [0,1] to replace the zero spectral counts. An iterative process was used where zero spectral counts were replaced by a fraction of a spectral count between 0 and 1, and the normality of the resulting ln(dNSAF) distribution was evaluated by the Shapiro–Wilk test. It is advisable to use the smallest value in order not to change the total sum significantly. An in-house program written in Java was used to extract peptides from the database results obtained from Mascot Percolator, select quantifiable proteins and then calculate spectral counts and dNSAF values. R program is used to evaluate the normality of ln(dNSAF) distribution by Shapiro–Wilk test. p-Value > 0.05 indicates the distribution can be considered as Gaussian distribution.

2.8. Sample preparation and protein quantification of standard proteins

To validate our use of Mascot Percolator results in label-free spectral counting quantitation and dNSAF normalization we used an NKL cell lysate spiked with known amounts of commercially available proteins from other species. For this analysis NKL cell lysates were obtained from 10 million NKL cells using the same method as described in the "Sample collection" section. Both standard proteins and NKL cell lysates were digested by trypsin using the same protocol as described in the "Proteolysis" section. The amounts of tryptic digests of each protein standard spiked into 10 μg tryptic digests of NKL cell lysates were 4 μg, 1 μg, 0.25 μg, 62.5 ng, 15.625 ng, 3.91 ng of BSA, bCYC and eMYG, respectively. Each sample (300 ng) was injected onto a Waters nanoAcquity UPLC (Waters Corp., Milford, MA) with an amaZon ion trap mass spectrometer (Bruker Daltonics, Billerica, MA). Each sample was run in triplicate. For the chromatographic separation, solvent A consisted of 0.1% formic acid in water and solvent B consisted of 0.1% formic acid in ACN. 5 μL of each sample was injected onto a Waters Symmetry C18 5 μm 180 μm × 20 mm precolumn at a flow rate of 5 μL/min for 5 min at 95% A/5% B, followed by peptide separation performed on Waters BEH130 1.7 μm C18 100 μm × 100 mm analytical column using gradient from 0 to 45% solvent B at 300 nL/min over 120 min parameters for the acquisition of precursor ions and MS/MS spectra were the same as that of described in the "LC–ESI ion trap mass spectrometry and MS/MS analysis" section. The methods for database search and protein quantification were the same as described above.

2.9. Real time PCR

RT-PCR was performed to evaluate levels of mRNAs for some of the proteins we identified as being differentially expressed in our proteomic approach. The untreated and IL-2 treated NK cells were homogenized in Trizol (Sigma, Cat No. T9424) and

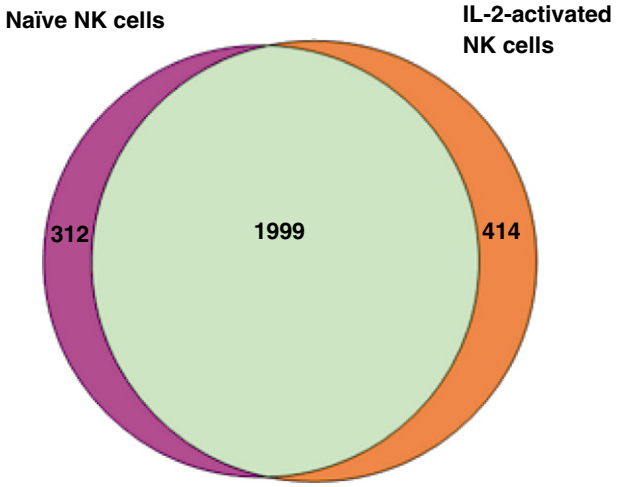


Fig. 2 – Venn diagram depicting the total number of proteins identified in naïve or IL-2-activated NK cells. Human primary NK cells isolated from three healthy donors were cultured in medium supplemented with or without IL-2 for 16 h. Total proteins were extracted and digested by trypsin follow by 2D LC–MS/MS analysis. 2311 proteins were identified from naïve NK cell samples, while 2413 proteins were identified in IL-2-activated NK cells. 1999 proteins were commonly identified in both conditions.

RNA was extracted according to the manufacturer's instructions. The RNA was reversely transcribed into cDNA using Omniscript RT kit from Qiagen (Cat. No. 205111). Quantitative real-time PCR was performed using the SYBR green chemistry (SsoFast Evagreen Supermix from BioRad, Cat. No. 172-5201) in a CFX96 real-time PCR detection system. The qPCR validated primers used to amplify PTP1B, CD97, and PCNA were obtained from SA Biosciences (real time PCR primers for Human PTP1B, Cat No. PPH00730C; CD97, Cat No. PPH07186A; PCNA, Cat No. PPH00216B and S27, Cat No. PPH17248B). The three step cycling conditions used were, 95 °C for 30 s, followed by 40 cycles of the denaturation at 95 °C for 1 s, and annealing at 60 °C for 5 s. The $2^{-\Delta\Delta CT}$ method was used for relative quantitation of gene expression. Data was analyzed using the BioRad CFX manager and GraphPad Prizm.

2.10. *Flow cytometry*

Flow cytometry was employed to determine purity of the isolated NK cells (defined $CD3^-/CD16^+/CD56^+$ cells), monitor their activation status after treatment with IL-2 by determining the expression level of CD69, and validate the differential expression CD48, CD56, CD11b, and CD11c. To determine purity NK cells were labeled with anti-CD3, -CD16, and -CD56 antibodies conjugated with PE, FITC, and APC, respectively, as described in our previous work [28,29]. CD69, CD48, CD11b, and CD11c antibodies used in the study were conjugated to either FITC or PE as denoted in the figures. Following labeling with the antibodies, the NK cells were washed and analyzed on LSR-II (BD Biosciences) flow cytometer. FlowJo® software was used to analyze the data.

3. Results and discussion

3.1. *Identification and quantification of proteins in naïve and IL-2-activated NK cells*

In recent years, targeted MS-based approaches have been employed to analyze the proteome of human NK-like cell lines, YTS and NKL [22–26,34]. NK-like cell lines and primary NK cells differ, substantially in the expression of cell surface receptors and intracellular signal transduction proteins [34]. Therefore, primary human NK cells were predominantly used in the current study. We anticipated that the complexity of unfractionated cell lysates would make it difficult to identify less abundant proteins. We therefore employed an orthogonal chromatographic step to reduce sample complexity prior to MS analysis to improve overall protein identifications. We adopted a two-dimensional separation approach utilizing

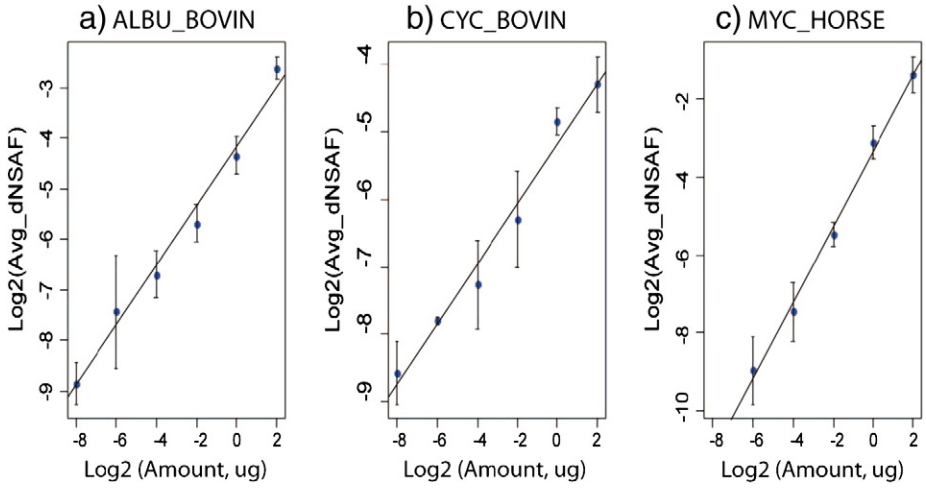


Fig. 3 – Linear regression between dNSAF values and known amount of protein standards BSA (a), cytochrome C (b) and myoglobin (c). Log2-transformed dNSAF values were plotted as a function of log2-transformed protein amounts in micrograms.

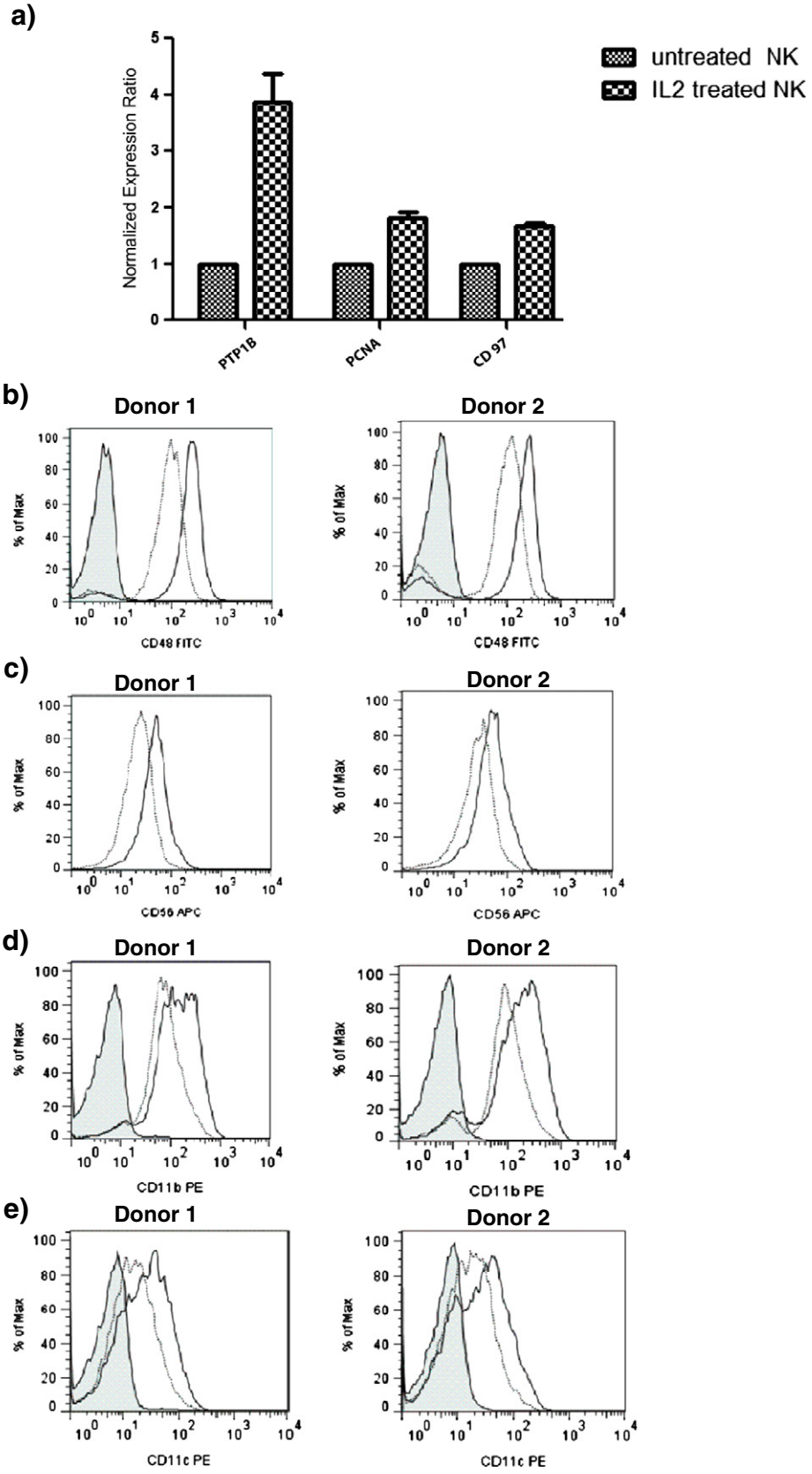


Fig. 4 – Validation of selected factors differentially expressed in IL-2 stimulated NK cells. Real-time-PCR analysis (a) of mRNA level of PTP1B, PCNA, and CD97 in naïve NK cells and NK cells activated by IL-2 for 16 h. Flow cytometry was used to monitor differential expression of CD48 (b), CD56 (c), CD11b (d), and CD11c (e) on naïve NK cells (dotted line) and NK cells stimulated with IL-2 (solid line) for 16 h. For each factor, data obtained for NK cells isolated from two healthy donors is shown. Non-specific IgG control is (shaded histogram) is shown only for the CD48 data.

high-pH RPLC as the first dimension separation. Increasing the number of fractions collected during first dimension liquid chromatography typically results in better separation that allows improved protein identification by downstream LC–MS/MS analysis [35]. However, since greater number of fractions also results in increased overall analysis time it is important to minimize the numbers of fractions but still produce the best output for LC–MS/MS analysis. To address the limitation of fraction collection but at the same time improve the orthogonality of the 2D separation, we adopted a novel RP-RPLC approach reported by Zou and colleagues [30]. For this approach, 20 fractions were collected in the first dimensional RPLC. Every two fractions with equal collection time interval were pooled, one from the early eluted section and the other from a later elution section (fractions 1 and 11, 2 and 12, and so on). Through this combination only half the number of fractions was submitted to LC–MS/MS analysis, significantly reducing the overall analysis time.

For the present work we applied Mascot Percolator which was embedded into the Mascot search engine to improve peptide and protein identifications. We compared the number of peptide spectrum matches (PSMs) identified with Mascot and Mascot Percolator at FDR = 0.1% and it was shown that more peptides could be identified by Mascot Percolator (data not shown). In total, 2311 proteins (≥1 unique peptide) were identified from three naïve NK cell samples, while 2413 proteins were identified from their IL-2 stimulated counterparts. The complete list of identified proteins from each donor is shown in the Supporting information (Table S1). Fig. 2 shows the Venn Diagrams of the numbers of proteins identified from naïve or IL-2 activated NK cells isolated from three healthy donors. There were 1999 proteins in common that were identified in both cells, whereas 312 proteins and 414 proteins were identified in only naïve or IL-2 activated NK cell, respectively.

To assess differences between protein abundance in naïve NK cells and IL-2-activated NK cells, spectral counting, a label-free protein quantification method was used. To account for shared peptide sequences among protein isoforms and achieve better accuracy, we adopted the distributive normalized spectral abundance factor (dNSAF) strategy previously reported by Zhang et al. [32]. However, given that the application of Mascot Percolater in peptide and protein identification may affect the performance of this approach, we first confirmed its effectiveness and reproducibility by spiking known amounts of protein standards into complex mixture of NKL cell lysates. To estimate the dynamic range and determine whether there was a linear correlation between the known amount of protein and their measured dNSAF values, the tryptic digests of BSA, bCYC and eMYG were spiked into tryptic digests of NKL cell lysates where the amounts of three standard proteins distributed evenly over 3 orders of magnitude in logarithmic scale. After protein identification with Mascot Percolator, dNSAF values for all quantifiable proteins were calculated. dNSAF values for BSA, bCYC and eMYG in different samples were extracted. Fig. 3 shows the linear regression between dNSAF values and known protein amounts where log2-transformed dNSAF values were plotted as a function of log2-transformed protein amounts in micrograms. The results demonstrated acceptable performance as there was a linear correlation coefficient of >0.990 for all three protein standards (Fig. 3a–c). The linear dynamic range for BSA and bCYC are three orders of magnitude (1024) while the dynamic range for eMYG is two orders of magnitude (256), which is probably due to the poor detection at lower concentration, as the presence of heme stabilizes the structure of eMYG and makes it resistant to digestion [36]. Based on this result, we concluded that the application of Mascot Percolater for peptide and protein identification did not affect the accuracy of protein quantification by spectral counting used in the present work. The linearity of our quantification method is maintained between protein amounts and dNSAF values over a dynamic range of at least three orders of magnitudes.

We analyzed the proteomes from three donors (three biological replicates) where each biological replicate contained three technical replicates to account for the biological and technical variation across datasets prior to protein quantification. According to the criteria that a protein is quantifiable if it can be detected in at least two of three technical replicates in all three biological replicates, 1375 proteins identified from both naïve NK cells and IL-2 activated NK cells were selected for quantification analysis. The dNSAF value of each protein was calculated and compared between the two conditions (naïve vs. IL-2-activated NK cells). Fold-change was calculated as the ratio of dNSAF of protein in IL-2-activated NK cells over that of naïve NK cells. Threshold levels for significantly up- or down-regulated proteins were set to more than 2-fold or less than 0.5-fold with $p \leq 0.05$ from Student *t*-test. Altogether, 436, 420 and 456 proteins exhibited significant abundance differences after IL-2 activation in NK cells isolated from the three donors, respectively. The complete list of up- and down-regulated proteins from each donor is shown in the Supporting information (Table S2). An overlap of 383 proteins was observed across all three donors and demonstrated similar trend of up or down-regulation (Table S2). While 301 proteins were up-regulated following IL-2 activation, 82 proteins were down-regulated (Table S2). In the discussion that follows, we focus on a select group of factors that, according to our proteomic analysis, are differentially expressed in the IL-2 stimulated human NK cells but have not been extensively studied in the IL-2-mediated modulation of NK cell responses.

3.2. *Functional annotation of IL-2-regulated proteins identified by quantitative analysis*

3.2.1. *Activation of JAK–STAT pathway and cell proliferation*

IL-2 critically regulates the proliferation and cytotoxicity of human NK cells. It is well known that IL-2 mediates its effects through the activation of the JAK–STAT pathway in which STAT1, STAT3 and STAT5 are activated. In our analysis, three STAT molecules STAT1, STAT3 and STAT4 were found to be up-regulated upon IL-2 stimulation. While the participation and the role of STAT1 and STAT3 in IL-2 signaling have been well documented, the function of STAT4 is still under investigation. Wang et al. reported IL-2 induced STAT4 activation as an alternative to the well-established JAK–STAT pathway in primary NK cells but not in T cells [37], which may explain why IL-2 enhances cytoxicity in NK cells but not in T cells, even though both cells have identical JAK–STAT signaling pathway. Moreover, they investigated the effect of IL-2 on IL-12-activated signaling pathways in NK cells

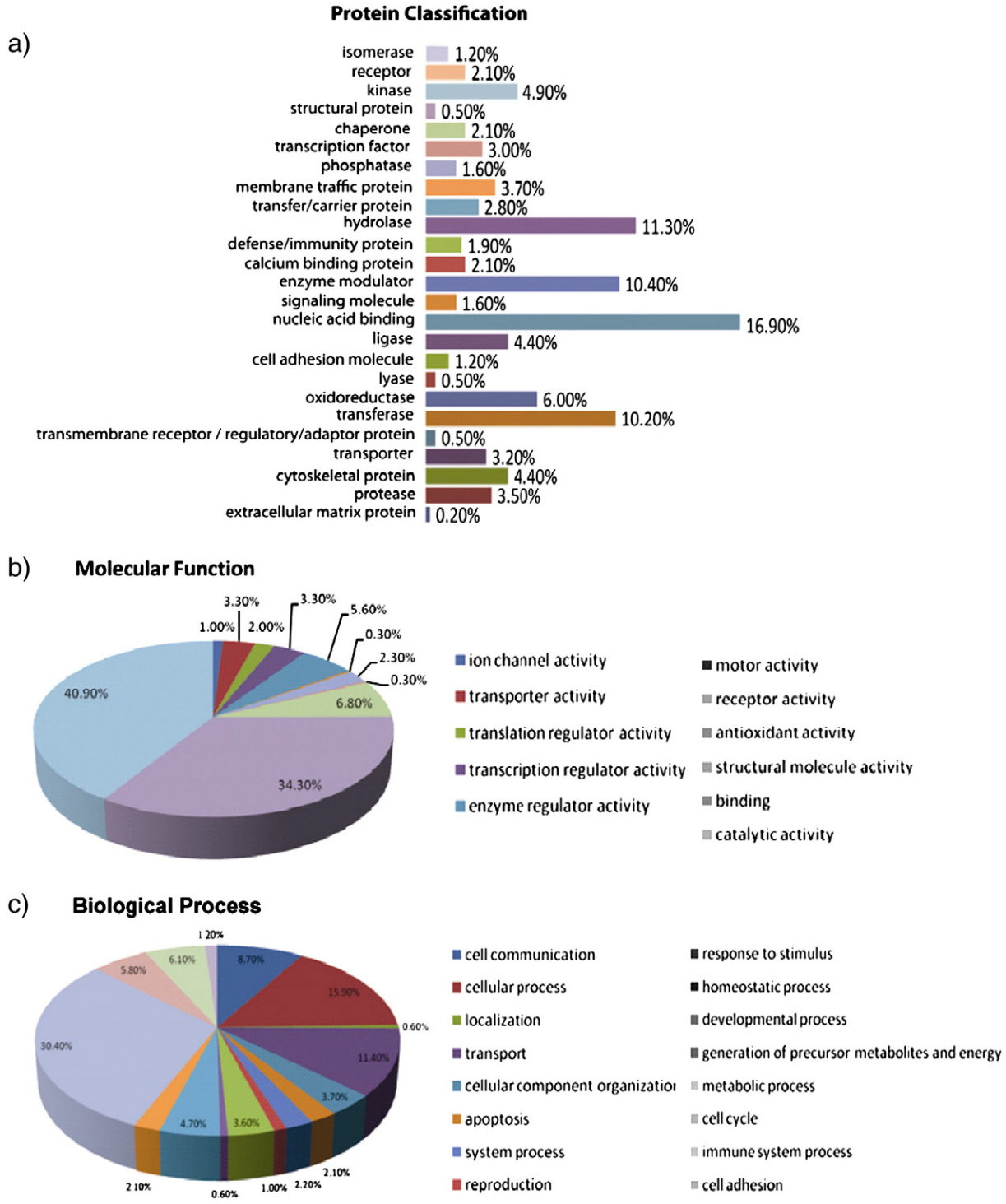


Fig. 5 – Gene ontology analysis of up- and down-regulated proteins. (a) Protein functional classification. 383 proteins that were found differentially expressed across all three donors could be classified into 25 categories, of which the top four are nucleic acid binding proteins (17%), hydrolase (11.3%), enzyme modulator (10.4%), and transferase (10.2%). (b) Molecular function. More than 70% of IL-2 regulated proteins have molecular function related to catalytic activities (41%) and binding (34%). (c) Biological process. IL-2 regulated proteins were found to be involved in 16 biological processes, of which the top three categories are: metabolic process (30.4%), cell process (15.9%) and transport (11.4%). Gene ontology analysis was performed in PANTHER database (http://www.pantherdb.org/).

and demonstrated that pretreatment of IL-2 promoted expression of high level of STAT4 through which the response of NK cells to IL-12 was enhanced [38]. Given that the results of current clinical trials for immunotherapy using IL-2 or IL-12 alone were not shown to be quite as successful as expected [39,40], STAT4 may prove to be a target for the study of

synergistic effect between IL-2 and IL-12 for developing more effective immunomodulatory strategies. Another important protein involved in JAK–STAT signaling pathway is protein-tyrosine phosphatase 1B (PTP1B), which was also shown to be up-regulated by IL-2 and confirmed by RT-PCR (Fig. 4a). PTP1B is an important phosphatase that plays both negative and positive roles in diverse signaling pathways. Although it has been extensively studied as a negative regulator of insulin and leptin signaling, and more recently as a positive factor in tumorigenesis, the role of PTP1B during immune cell signaling is not well characterized. PTP1B and T-cell PTP (TC-PTP) form the first non-transmembrane subfamily of PTPs [41]. While previous studies have identified JAK1 and JAK3 as TC-PTP substrates and implicated TC-PTP in the regulation of JAK–STAT signaling activated by IL-2, PTP1B was unable to bind JAK1 and JAK3 [42]. However, other studies demonstrated that after IFN stimulation PTP1B targets two other JAK family members, JAK2 and tyrosine kinase 2 (TYK2) [43]. Also, PTP1B has been implicated in the dephosphorylation of STAT5 in prolactin signaling [44]. Although PTP1B does not appear to be directly associated with the regulation of JAK–STAT activation, it might play an important role in IL-2 signaling through another mechanism.

As expected, many proteins identified to be up or down-regulated are involved in DNA replication, translation initiation/elongation/termination or the regulation of cell cycle since one of the known results of NK cell exposure to IL-2 is stimulation of cell proliferation. Three DNA replication licensing factor MCM2, MCM5 and MCM7 were found to be up-regulated. DNA replication licensing is the process of ensuring that chromosomes are duplicated once and only once during the cell cycle. The MCM proteins are required for DNA replication licensing and consist of a group of ten conserved factors functioning in the replication of the genomes of archae and eukaryotic organisms. Among these, MCM2–7 proteins are related to each other and form a family of DNA helicases at the initiation step of DNA synthesis. While the other MCM proteins were not detected in our analysis possibly due to their low abundance, the elevated expression of MCM2, MCM5 and MCM7 implicated that increased NK cell proliferation may be attributable to these proteins. Lymphocyte proliferation is often used to mark immune response following immunotherapy. Identifying proteins that improve monitoring of NK cell proliferation to determine the extent of the immune response may prove to be crucial for clinical decision-making during the immunotherapy in order to minimize the risk of toxicity in patients.

Another protein that is associated with the proliferation process of NK cells is proliferating cell nuclear antigen (PCNA) which is synthesized in early G1 and S phases of the cell cycle. Increased PCNA expression was also observed in our proteomic analysis. We validated our proteomic observations using real time PCR and confirmed that the level of PCNA mRNA expression was increased by 2-fold in NK cells after 16 h of stimulation by IL-2 (Fig. 4a). PCNA is a marker that can be used to detect early stage T-cell proliferation, as it was found that unstimulated human peripheral blood T-lymphocytes were PCNA negative and expression of PCNA in these cells increased after stimulation [45]. In the clinical setting, PCNA can be used as a marker to monitor T-cell function that reflects immune condition in patients undergoing immunosuppressive therapy [46,47]. Such monitoring of PCNA would be crucial in reducing the risks and providing optimal immunosuppressive therapy for each transplant recipient [46–48]. Recently, PCNA mRNA level in the peripheral blood were measured by real-time RT-PCR to monitor patient's immune condition after renal transplantation [49]. Increased PCNA expression observed in our analysis suggest that monitoring the levels of this antigen may provide useful guidance of the NK cell proliferation status in patients treated with immunotherapies. PCNA is recruited to the NK cell immune synapse and via its interaction with the NKp44 receptor attenuates NK cell function. In this context PCNA expression allows cancer cells to escape NK cell attack. It remains to be determined if the increased expression of PCNA represents a compensatory mechanism to attenuate immunological function of the IL-2-stimulated NK cells [50,51].

3.2.2. *Cluster of differentiation (CD) molecules*

In addition to changes in molecules that regulate cell proliferation and activation of cytotoxicity in NK cells, changes in cluster of differentiation (CD) molecules were also expected, as their functions are often closely tied to the immune system. While CD56, CD48, CD98, CD97, CD225, and CD300a showed increased expression in our analysis, notably, there were decreases in the levels of CD11b, CD11d, CD11c and CD43 after IL-2 stimulation. Flow cytometry assays were used to validate the increased expression of CD48 and CD56 on the IL-2 activated NK cells (Fig. 4b, c). CD48 is a glycosyl-phosphatidyl-inositol (GPI)-anchored protein expressed on the surface of NK cells and is known as a co-stimulatory factor and a high affinity ligand for natural killer cell receptor 2B4 [52,53]. 2B4/CD48 interactions in NK cells are required for the enhanced proliferation and the development of optimal cytolytic and secretory NK effector functions during IL-2 activation [54]. Increased expression of CD48 on the NK cells may allow this ligand to be presented to the 2B4 receptor on opposing NK, T or B cells, thereby priming the NK cells to express IL-13 in the presence of another co-stimulatory signal [55].

Human CD97 is a member of the EGF-TM7 family of adhesion class heptahelical receptors and was identified as an early activation marker for human lymphocytes [56]. We observed and validated increased CD97 level in response to IL-2 by real time RT-PCR (Fig. 4a) that was comparable with the results from Kop et al. [57].

CD98 is a transmembrane glycoprotein identified as a lymphocyte activation antigen [58,59]. Because the level of CD98 on cell surface was markedly increased in activated lymphocytes, CD98 has been mainly used as a T cell activation marker [58,59] and has been implicated for its role in regulating integrin signaling, amino acid transport and immune response [60,61]. Integrin–CD98 interaction acts as a co-stimulatory signal in T cells [62,63]. So far no studies have investigated the role of CD98 in NK cell activation.

CD300a is a cell surface inhibitory receptor that is expressed in all human NK cells and is known to down-regulate the cytotoxicity of NK cells [64]. Different studies have also evaluated the CD300a activity in immune regulation in other immune cells such as plasmacytoid dendritic cells [65], T cells [66] and neutrophils [67]. However, little is known about the functionality and the ligand of CD300a in NK cells so far and it was only recently that Nakahashi-Oda et al. identified

phosphatidylserine (PS), which is exposed on the outer leaflet of the plasma membrane of apoptotic cells, as a ligand for CD300a [68]. Alvarez et al. reported that inflammatory stimuli such as granulocyte macrophage-colony stimulating factor (GM-CSF) or lipopolysaccharide (LPS) could induce a significant increase in cell surface expression of CD300a in neutrophils and that the signaling through this receptor down-regulated neutrophil function [67]. Activation and regulation of CD300a by cytokines in NK cells have never been reported but our results and the research in other cell-types suggest that CD300a may play a critical role in NK cell responses. Further investigation is required to understand how CD300a is regulated by IL-2 and the physiological role of CD300a in NK cells.

CD225 is also known as interferon (IFN)-induced transmembrane protein 1 (IFITM1) which is a member of the IFN-inducible transmembrane protein family [69]. It is known that the transcription of CD225 is induced by IFN-γ [70], and that CD225 can mediate IFN-γ-induced inhibition of cell proliferation and inhibit ERK activation [71]. CD225 has also been implicated in the control of cell growth, as it can arrest cell cycle progression in the G1 phase in a p53-dependent manner [71]. The role of CD225 in NK cell activation has never been reported. However, as an important factor for growth control, it is likely that CD225 mediates the negative regulation and plays an anti-proliferative role during NK cell activation.

Surprisingly, proteomic analysis indicated that IL-2 stimulation results in decreased expression of the integrin α subunits CD11b, CD11c and CD11d. Even CD43, which along with the CD11 molecules are described as NK cell maturation markers [72]. Flow cytometry analysis indicated that IL-2 stimulation results in increased expression of CD11b and CD11c on the NK cell surface (Fig. 4d, e). We are currently investigating if methodological issues resulting from the processing of the NK cell lysates or the generation of the tryptic peptides from these lysates contribute to the discrepancy in these results.

3.3. *Gene ontology and pathway analysis*

To obtain more information about the molecular function of IL-2 regulated proteins and to identify pathways possibly involved in IL-2 activation in NK cells, we performed gene ontology and pathway analysis using the PANTHER database (http://www.pantherdb.org/). Overall, 383 proteins that were found differentially expressed across all three donors could be classified into 25 categories of which the top four are nucleic acid binding proteins (17%), hydrolases (11.3%), enzyme modulators (10.4%), and transferases (10.2%) (Fig. 5a). This is in agreement with the fact that the activation of NK cells in response to cytokines requires protein regulators to bind to DNA and activate new protein synthesis for cell proliferation and enhanced cytotoxicity. An analysis of the molecular function of these proteins revealed that most of these proteins are involved in catalytic activities (41%) and binding (34%) (Fig. 5b). Furthermore, these proteins were found to be involved in various biological processes, of which the top three categories are: metabolic process (30.4%), cell process (15.9%) and transport (11.4%) (Fig. 5c). Taken together, the data from gene ontology analysis suggested that most of these proteins are involved in critical cellular events triggered by IL-2 in the NK cells.

In search for novel pathways that might be involved in IL-2 signaling in NK cells, a pathway analysis was also performed in PANTHER database and a total of 90 pathways were found to be related to the IL-2 regulated proteins based on this study. In addition to the previously known pathways, such as JAK–STAT pathway, MAPK/ERK and NF-κB pathway, other pathways that are very likely to be associated with IL-2 signaling include: ubiquitin proteasome pathway, integrin signaling pathway, PDGF signaling pathway, EGFR signaling pathway and Wnt signaling pathway. Table 1 provides the names of IL-2 regulated proteins identified in our analysis that are known to be the components of these pathways. Although the

Table 1 – Novel pathways that may be involved in IL-2 signaling in human primary NK cells and names of IL-2 regulated proteins that are known to be the components of these pathways.

Ubiquitin proteasome pathway (11)
26S proteasome non-ATPase regulatory subunit 1 (PSMD1)
26S proteasome non-ATPase regulatory subunit 3 (PSMD3)
26S proteasome non-ATPase regulatory subunit 4 (PSMD4)
26S proteasome non-ATPase regulatory subunit 7 (PSMD7)
26S proteasome non-ATPase regulatory subunit 11 (PSMD11)
26S proteasome non-ATPase regulatory subunit 12 (PSMD12)
NEDD8-activating enzyme E1 catalytic subunit (UBA3)
26S protease regulatory subunit 4 (PRS4)
Ubiquitin-conjugating enzyme E2 L3 (UB2L3)
Ubiquitin carboxyl-terminal hydrolase isozyme L5 (UBP5)
SUMO-activating enzyme subunit 1 (SAE1)
Integrin signaling pathway (9)
Integrin beta-7 (ITB7)
Dual specificity mitogen-activated protein kinase kinase 1 (MP2K1)
Integrin alpha-X (ITAX)
Ras-related C3 botulinum toxin substrate 3 (RAC3)
ADP-ribosylation factor 3 (ARF3)
Rho-related GTP-binding protein RhoB (RHOB)
Actin, gamma-enteric smooth muscle (ACTH)
Proto-oncogene tyrosine-protein kinase Fyn (FYN)
Integrin alpha-D (ITAD)
PDGF signaling pathway (8)
Dual specificity mitogen-activated protein kinase kinase 1 (MP2K1)
Signal transducer and activator of transcription 1-alpha/beta (STAT1)
Signal transducer and activator of transcription 3 (STAT3)
Signal transducer and activator of transcription 4 (STAT4)
Ras-related protein Rab-11B (RB11B)
Rho GTPase-activating protein 4 (RHG04)
Rho-related GTP-binding protein RhoB (RHOB)
Ribosomal protein S6 kinase alpha-3 (KS6A3)
EGF receptor signaling pathway (7)
Dual specificity mitogen-activated protein kinase kinase 1 (MP2K1)
Signal transducer and activator of transcription 1-alpha/beta (STAT1)
Signal transducer and activator of transcription 3 (STAT3)
Signal transducer and activator of transcription 4 (STAT4)
Ras-related C3 botulinum toxin substrate 3 (RAC3)
Serine/threonine-protein phosphatase 2A catalytic subunit beta isoform (PP2AB)
Mitogen-activated protein kinase 14 (MK14)
Wnt signaling pathway (6)
Beta-arrestin-1 (ARRB1)
Serine/threonine-protein phosphatase 2A catalytic subunit beta isoform (PP2AB)
Histone deacetylase 1 (HDAC1)
C-terminal-binding protein 2 (CTBP2)
Nuclear factor of activated T-cells, cytoplasmic 2 (NFAC2)
Actin, gamma-enteric smooth muscle (ACTH)

roles of these pathways in promoting or regulating the function of NK cells have been investigated in the past, so far none of them have been directly linked to the activation of NK cells by IL-2, and therefore could be the new targets for future studies. Given the pleiotropic effects of IL-2 and diverse functions of IL-2-activated NK cells, there should be many possible downstream effectors that may be necessary to drive IL-2-induced effects.

4. Conclusions

In summary, we have developed an effective method for the proteomic analysis in human primary NK cells. We successfully employed 2D LC to reduce the sample complexity prior to MS analysis. To improve protein identification, Mascot Percolator was employed, with 2311 and 2413 proteins being identified from naïve and IL-2-activated NK cells, respectively. Label-free quantitative analysis via spectral counting revealed a list of 383 proteins that were either up or down-regulated in IL-2 signaling. Functional annotation of IL-2 regulated proteins in the present work revealed several proteins with important functions related to IL-2 signaling that could potentially serve as targets for future investigation of IL-2 signaling in human primary NK cells. A pathway analysis was also performed and revealed several novel pathways not previously known to be involved in IL-2 signaling. The quantitative proteomic analysis in present work provided a comprehensive view of proteins that may be associated with IL-2 signaling. Further functional analysis of proteins of interests will improve our understanding of signaling transduction and biological processes involved in NK cell activation by IL-2.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.jprot.2013.06.024.

Acknowledgments

This work is supported by the Department of Defense Pilot Award Grant (W81XWH-11-1-0181). LL acknowledges an H. I. Romnes Faculty Research Fellowship.

REFERENCES

[1] Trinchieri G. Biology of natural killer cells. Adv Immunol 1989;47:187–376.
[2] Moretta A, Bottino C, Vitale M, Pende D, Cantoni C, Mingari MC, et al. Activating receptors and coreceptors involved in human natural killer cell-mediated cytolysis. Annu Rev Immunol 2001;19:197–223.
[3] Smith KA. Interleukin-2: inception, impact, and implications. Science 1988;240:1169–76.
[4] Caligiuri MA, Zmuidzinas A, Manley TJ, Levine H, Smith KA, Ritz J. Functional consequences of interleukin 2 receptor expression on resting human lymphocytes. Identification of a novel natural killer cell subset with high affinity receptors. J Exp Med 1990;171:1509–26.
[5] Bonnema JD, Rivlin KA, Ting AT, Schoon RA, Abraham RT, Leibson PJ. Cytokine-enhanced NK cell-mediated cytotoxicity. Positive modulatory effects of IL-2 and IL-12 on stimulus-dependent granule exocytosis. J Immunol 1994;152:2098–104.
[6] Karnitz LM, Abraham RT. Interleukin-2 receptor signaling mechanisms. Adv Immunol 1996;61:147–99.
[7] Russell SM, Johnston JA, Noguchi M, Kawamura M, Bacon CM, Friedmann M, et al. Interaction of IL-2R beta and gamma c chains with Jak1 and Jak3: implications for XSCID and XCID. Science 1994;266:1042–5.
[8] Ellery JM, Nicholls PJ. Alternate signalling pathways from the interleukin-2 receptor. Cytokine Growth Factor Rev 2002;13:27–40.
[9] Johnston JA, Kawamura M, Kirken RA, Chen YQ, Blake TB, Shibuya K, et al. Phosphorylation and activation of the Jak-3 Janus kinase in response to interleukin-2. Nature 1994;370:151–3.
[10] Nelson BH, Willerford DM. Biology of the interleukin-2 receptor. Adv Immunol 1998;70:1–81.
[11] Frank DA, Robertson MJ, Bonni A, Ritz J, Greenberg ME. Interleukin 2 signaling involves the phosphorylation of Stat proteins. Proc Natl Acad Sci U S A 1995;92:7779–83.
[12] Garcia-Lora A, Martinez M, Pedrinaci S, Garrido F. Different regulation of PKC isoenzymes and MAPK by PSK and IL-2 in the proliferative and cytotoxic activities of the NKL human natural killer cell line. Cancer Immunol Immunother 2003;52:59–64.
[13] Yu TK, Caudell EG, Smid C, Grimm EA. IL-2 activation of NK cells: involvement of MKK1/2/ERK but not p38 kinase pathway. J Immunol 2000;164:6244–51.
[14] Zhou J, Zhang J, Lichtenheld MG, Meadows GG. A role for NF-kappa B activation in perforin expression of NK cells upon IL-2 receptor signaling. J Immunol 2002;169:1319–25.
[15] Ussat S, Scherer G, Fazio J, Beetz S, Kabelitz D, Adam-Klages S. Human NK cells require caspases for activation-induced proliferation and cytokine release but not for cytotoxicity. Scand J Immunol 2010;72:388–95.
[16] Smith FO, Downey SG, Klapper JA, Yang JC, Sherry RM, Royal RE, et al. Treatment of metastatic melanoma using interleukin-2 alone or in conjunction with vaccines. Clin Cancer Res 2008;14:5610–8.
[17] Klapper JA, Downey SG, Smith FO, Yang JC, Hughes MS, Kammula US, et al. High-dose interleukin-2 for the treatment of metastatic renal cell carcinoma: a retrospective analysis of response and survival in patients treated in the surgery branch at the National Cancer Institute between 1986 and 2006. Cancer 2008;113:293–301.
[18] Buhtoiarov IN, Neal ZC, Gan J, Buhtoiarova TN, Patankar MS, Gubbels JA, et al. Differential internalization of hu14.18-IL2 immunocytokine by NK and tumor cell: impact on conjugation, cytotoxicity, and targeting. J Leukoc Biol 2011;89:625–38.
[19] Delgado DC, Hank JA, Kolesar J, Lorentzen D, Gan J, Seo S, et al. Genotypes of NK cell KIR receptors, their ligands, and Fc gamma receptors in the response of neuroblastoma patients to Hu14.18-IL2 immunotherapy. Cancer Res 2010;70:9554–61.
[20] Gubbels JA, Gadbaw B, Buhtoiarov IN, Horibata S, Kapur AK, Patel D, et al. Ab-IL2 fusion proteins mediate NK cell immune synapse formation by polarizing CD25 to the target cell-effector cell interface. Cancer Immunol Immunother 2012;60:1789–800.
[21] Yang RK, Kalogriopoulos NA, Rakhmilevich AL, Ranheim EA, Seo S, Kim K, et al. Intratumoral hu14.18-IL-2 (IC) induces local and systemic antitumor effects that involve both activated T and NK cells as well as enhanced IC retention. J Immunol 2012;189:2656–64.
[22] Lund TC, Anderson LB, McCullar V, Higgins L, Yun GH, Grzywacz B, et al. iTRAQ is a useful method to screen for membrane-bound proteins differentially expressed in human natural killer cell types. J Proteome Res 2007;6:644–53.
[23] Man P, Novak P, Cebecauer M, Horvath O, Fiserova A, Havlicek V, et al. Mass spectrometric analysis of the

Published in final edited form as:
J Proteomics. 2011 December 21; 75(2): 480–490. doi:10.1016/j.jprot.2011.08.013.

RT-SVR+q: A Strategy for Post-Mascot Analysis Using Retention Time and q Value Metric to Improve Peptide and Protein Identifications

Weifeng Cao[1], **Di Ma**[2], **Arvinder Kapur**[3], **Manish S Patankar**[3], **Yadi Ma**[4], and **Lingjun Li**[1,2,*]
Weifeng Cao: wcao2@wisc.edu; Lingjun Li: lli@pharmacy.wisc.edu
[1]Department of Chemistry, University of Wisconsin-Madison, 777 Highland Ave., Madison, WI 53705
[2]School of Pharmacy, University of Wisconsin-Madison, 777 Highland Ave., Madison, WI 53705
[3]Department of Obstetrics and Gynecology, University of Wisconsin-Madison, 777 Highland Ave., Madison, WI 53705
[4]Department of Computer Sciences, University of Wisconsin-Madison, 777 Highland Ave., Madison, WI 53705

Abstract

Shotgun proteomics commonly utilizes database search like Mascot to identify proteins from tandem MS/MS spectra. False discovery rate (FDR) is often used to assess the confidence of peptide identifications. However, a widely accepted FDR of 1% sacrifices the sensitivity of peptide identification while improving the accuracy. This article details a machine learning approach combining retention time based support vector regressor (RT-SVR) with q value based statistical analysis to improve peptide and protein identifications with high sensitivity and accuracy. The use of confident peptide identifications as training examples and careful feature selection ensures high R values (>0.900) for all models. The application of RT-SVR model on Mascot results (p=0.10) increases the sensitivity of peptide identifications. q value, as a function of deviation between predicted and experimental RTs(Δ RT), is used to assess the significance of peptide identifications. We demonstrate that the peptide and protein identifications increase by up to 89.4% and 83.5%, respectively, for a specified q value of 0.01 when applying the method to proteomic analysis of the natural killer leukemia cell line (NKL). This study establishes an effective methodology and provides a platform for profiling confident proteomes in more relevant species as well as a future investigation of accurate protein quantification.

Keywords

tandem mass spectrometry; shotgun proteomics; database search; support vector regressor; retention time; q value; peptide identification; NKL cell

*To whom correspondence should be addressed. lli@pharmacy.wisc.edu.

SUPPORTING INFORMATION AVAILBALE

Published in final edited form as:
Am J Reprod Immunol. 2012 July ; 68(1): 28–37. doi:10.1111/j.1600-0897.2012.01113.x.

The mucin MUC16 (CA125) binds to NK cells and monocytes from peripheral blood of women with healthy pregnancy and preeclampsia

Chanel Tyler[*], **Arvinder Kapur**[*], **Mildred Felder**[*], **Jennifer A. Belisle**[*], **Christine Trautman**[*], **Jennifer A.A. Gubbels**[†], **Joseph P. Connor**[*], and **Manish S. Patankar**[*,1]

[*]Department of Obstetrics and Gynecology, University of Wisconsin-Madison, Madison, WI

[†]Department of Biology, Augustana College, Sioux Falls, SD

Abstract

Problem—MUC16 (CA125) released from ovarian tumors binds to NK cells and monocytes via the inhibitory receptor Siglec-9. Here, we investigate if MUC16 also binds to circulating immune cells during pregnancy and in women with preeclampsia.

Method of study—MUC16 binding was monitored by flow cytometry and immunoprecipitation and RT-PCR was used to monitor indigenous expression in immune cells. Serum CA125 levels were measured by a clinical assay.

Results—MUC16 was equally distributed on Siglec-9^{pos} CD16^{pos}/CD56^{dim} and CD16^{neg}/CD56^{br} NK cells in the healthy pregnant and preeclampsia groups. While serum CA125 levels and number of NK and monocytes were similar, increased binding of MUC16 was observed on these immune cells in the preeclampsia cohort as compared to the healthy pregnant samples.

Conclusion—MUC16 binding to NK cells and monocytes likely contributes to tolerance of the fetal allograft from maternal responses and may also serve as a novel biomarker for preeclampsia.

Keywords

Immune cell subsets; MUC16; CA125; Siglec-9; Biomarker

Introduction

CA125, is a tumor biomarker used extensively to monitor epithelial ovarian cancer [1-3]. CA125 is a repeating peptide epitope present in the tandem repeat region of MUC16, a 3-5 million Da heavily glycosylated mucin overexpressed by epithelial ovarian tumors [4-6]. We have previously demonstrated that ovarian tumors utilize MUC16 to attenuate the cytolytic responses of NK cells [7, 8]. The large molecular weight and high negative charge of this heavily glycosylated mucin also acts as a barrier that prevents the NK cells from forming activating immunologic synapses with the ovarian tumor targets [6, 7, 9].

MUC16 is a membrane-spanning mucin that is initially expressed on the surface of epithelial cells and especially on the epithelial ovarian tumor cells [6, 10]. Proteolytic cleavage results in release of the mucin from the cell surface. The shed mucin, (sMUC16) molecules traverse to the peripheral blood from the extracellular milieu, where they can be detected using the

[1]Corresponding author: **Manish S. Patankar**, Department of Obstetrics and Gynecology, University of Wisconsin-Madison, H4/657 CSC, 600 Highland Avenue, Madison, WI-53792-6188; Tel: 608-262-8871; Fax: 608-265-6572; patankar@wisc.edu.

clinical serum CA125 test. In our analysis of peripheral blood mononuclear cells (PBMC) isolated from ovarian cancer patients and the immune cells isolated from their peritoneal fluid (PFMC) we observed that approximately 10-15% of B cells, 30-40% of NK cells and >90% of monocytes were positive for sMUC16 [11, 12]. Several lines of evidence indicate that the PBMC and PFMC do not express sMUC16 but instead specifically bind to the mucin released from ovarian tumors [11, 12].

We have now also demonstrated that sMUC16 predominantly binds to immune cells via Siglec-9 [11], a α2-3-linked sialic acid binding I-type lectin known to serve as an inhibitory immune cell receptor [13-17]. Siglec-9 is expressed on approximately 30-40% of the $CD16^{pos}$/$CD56^{dim}$ NK cell subsets and in ovarian cancer patients these cells are double positive for sMUC16 [11]. High expression of Siglec-9 is observed on >90% of the monocytes and a correspondingly high level of sMUC16 binding is observed on these immune cells in ovarian cancer patients.

MUC16 is expressed by endometrial epithelial cells and also in the decidua [18-21]. Indeed, serum CA125 levels increase during pregnancy [22], which is one of the reasons why serum CA125 cannot be used as an early diagnostic test exclusively for ovarian cancer. Considering our previous work on sMUC16 binding to PBMC and PFMC of ovarian cancer patients [11, 12], we investigated if the mucin is also present on specific subsets of immune cells of pregnant women.

Our results indicate that the binding pattern of sMUC16 to NK cells and monocytes from peripheral blood of pregnant women closely matches the expression of Siglec-9 on these immune cells. Important differences were observed in the subsets of NK cells from normal pregnant women and preeclampsia patients that bind to sMUC16 and express Siglec-9. Data presented in this study lays the groundwork for future studies on the potential biological significance of sMUC16 binding to immune cells in healthy pregnant women and preeclampsia patients. In addition, differences in the binding patterns of sMUC16 to NK cells and the expression of Siglec-9 on these cells may also be exploited for the development of a novel diagnostic test for the detection of preeclampsia.

Methods

Cell lines and reagents

Siglec-9 expressing Jurkat cells were a kind gift from Dr. Jim Paulson (Scripps Research Institute, Ca). The Jurkat cells were cultured in RPMI-1640 media supplemented with 10% fetal calf serum. All other reagents were commercially obtained. ECC-1 and OVCAR-3 cells were purchased from ATCC and were cultured in RPMI 1640 media containing 10% fetal Bovine serum. Fluorophore conjugated Anti-CD14 (PerCP-CY5.5, clone: M5E2), CD3 (APC-Cy7, clone SK7), CD56 (Alexa 700, clone B159), CD16 (PE-Cy7), clone 3G8), CD19 (PE, clone HIB19), Siglec-9 (CDw329; FITC, clone E10-286) were from BD Biosciences and secondary antibodies were purchased from Jackson ImmunoResearch. All other reagents were from Sigma or Invitrogen.

Subjects

All subjects signed an informed consent and the study was approved by the Institutional Review Boards of the University of Wisconsin-Madison and Meriter Hospital. The women were recruited at the time of admission to Labor and Delivery. Subjects were considered eligible controls if they had completed 37 weeks of gestation, had an uncomplicated prenatal course, and had no preexisting co-morbidities. Eligible preeclamptic subjects were identified using strict diagnostic criteria. Prenatal records were reviewed to ensure that there was no evidence of hypertension prior to twenty weeks of gestation. Hypertension was defined as

systolic blood pressures of at least 140 mm Hg or diastolic blood pressures of at least 90 mm Hg two times measured 6 hours apart. The definition of protienuria is the excretion of 300 mg of protein or greater in a 24 hour urine collection [23]. Urine dip analysis alone without confirmatory 24 hour urine protein excretion was not considered sufficient to enroll a subject into the study. Subjects were excluded from the study if they carried the diagnosis of chronic hypertension, diabetes, anti-phospholipid lipid antibody syndrome, or systemic lupus erythematous. Subjects were also excluded if given the diagnosis of abruption or had meconium stained amniotic fluid. Of the 26 women included in the study there was one African American and all others were Caucasian. Table 1 shows the baseline characteristics of the groups. The groups showed significant difference in age as well as gestational age which is expected when considering the disease process.

Isolation of peripheral blood lymphocytes

Blood samples were obtained upon delivery from full term uncomplicated pregnancies as well as pregnancies complicated by preeclampsia. Mononuclear cells were isolated under sterile conditions using Histopaque (Sigma Aldrich, St Louis, MO). The mononuclear layer was retained and washed once with PBS-BSA before being cryopreserved in 90% fetal bovine serum (FBS; Hyclone) containing 10% dimethyl sulphoxide (DMSO). Serum samples were layered over Histopaque with the mononuclear layer being isolated and washed and cryopreserved.

Analysis of Siglec-9 and MUC16 on NK cells by flow cytometry

Cryopreserved cells were thawed and washed (15 min at 300Xg in PBS-BSA). Incubation of cells with primary and secondary antibodies was performed for 30 min on ice. After incubation with each antibody, cells were washed with PBS-BSA at 300Xg for 10 min at 4°C. Cells were blocked with goat IgG prior to staining with anti-MUC16 antibody VK8. The cells were washed and Allophycocyanin (APC)-conjugated goat anti-mouse (GAM) IgG, antibody was added. The cells were then washed and incubated with mouse IgG for 20 min to bind any additional Fab sites on the GAM secondary. The cells were then incubated with a cocktail of fluorophore-conjugated antibodies to stain for CD3, CD45, CD56, CD16, and Siglec-9 [11, 12].

After the final wash, cells were resuspended in approximately 300 μl of PBS-BSA. Immediately prior to data acquisition on the LSRII flow cytometer (Beckton Dickinson, San Jose, CA), DAPI, a viability indicator was applied to each sample. Each day, quality control procedures were run on the LSRII prior to data acquisition. Beckton Dickinson Cytometry Setup and Tracking (CST) beads were run daily to reproducibly set up the cytometer. Use of these beads allows for determination of cytometer baseline, the standard deviation of electronic noise, assessment of photomultiplier tube (PMT) drift, and adjusts cytometer settings to maximize population resolution in each detector. These measurements also verify that data is acquired within the linear range of each PMT. Spherotech Rainbow Beads, both mid-range and ultra, were run each day to determine that for a given PMT voltage, nearly identical target values (mean fluorescence intensity (MFI), were obtained for each detector of interest. Setting the instrument up in this manner, with these controls, allowed for the negative controls from each patient for each day to give rise to comparable MFIs in each fluorescence channel. From here, compensation controls were generated and acquired for experiments done on individual days to adjust for spectral overlap between PMTs. Flow Jo software was used to analyze the flow cytometry data.

CA-125 Assay

Serum was separated and the number of units of CA125 in the sample was determined by using the standard clinical radioimmunoassay (Abbott Axsym, Abbott Park, IL). CA125

measurements were conducted in the clinical pathology laboratory of the University of Wisconsin-Madison Hospital and Clinics.

Immunoprecipitation of sMUC16 from immune cell lysates

The PBMC's from healthy donor, ovarian cancer patient and OVCAR-3 ovarian cancer cell line were lysed in RIPA buffer containing protease inhibitors. After calculating the protein concentration (BCA) lysate equivalent to 500 μg of total protein was mixed with 100 μg of VK-8 antibody and tubes were rotated overnight at 4°C. Protein G-agarose beads (100 μl) were added to the mixture and the tubes were rotated for 1 h at room temperature. Suspensions were centrifuged (400Xg for 1 min), supernatants were removed, and the beads were washed three times with RIPA buffer (500 μl) containing protease inhibitors. Beads were then resuspended in 30 μL Laemmli buffer, boiled for 5 min and loaded on 7.5% SDS-PAGE gel. Proteins from the gel were transferred to PVDF membranes and MUC16 was detected using the VK-8 as the primary antibody and a HRP-conjugated goat anti-mouse as the secondary antibody.

MUC16 RT-PCR

OVCAR-3, ECC-1, and PBMC samples were homogenized in Trizol and RNA was isolated according to the manufacturer's instructions (Invitrogen). The RNA was reverse transcribed into cDNA using Omniscript RT kit from Qiagen (Cat.No.205111). The MUC16 and GAPDH (Glyceraldehyde 3-phosphate dehydrogenase) were amplified with the respective primers (MUC16 forward: 5′GCCTCTACCTTAACGGTTACAATGAA3′; reverse: 5′GGTACCCCATGGCTGTTGTG-3:, and GAPDH forward: GAGTCAACGGATTTGGTCGT, reverse: TTGATTTTGGAGGGATCTCG) using fast cycling PCR kit from Qiagen as described in our earlier studies [12, 24]. MUC 16 was amplified using the initial activation for 10 minutes at 95°C, followed by 35 cycles of 15 seconds at 95°C and 1 minute at 60°C. The cycling conditions used for amplifying GAPDH were, initial activation at 95°C for 5min, followed by 30 cycles of the denaturation at 96°C for 5seconds, annealing at 60°C for 5seconds and extension at 68°C for 10seconds, with final extension at 72°C for 1min. GAPDH was used as an endogenous reference to determine the integrity of the mRNA in each sample. The PCR product was run on 2.5% agarose gel at 100V for 2 hr and bands were visualized using Flourchem8900 ultraviolet transilluminator.

Statistics

The Flo Jo flow cytometry software was used to obtain the raw data on the percentage of immune cells positive for sMUC16, Siglec-9 and other immune markers. The data was plotted using the Graph Pad statistical software. Statistical significance of the data was determined using the non-parametric Mann Whitney U test.

Results

sMUC16 binds to immune cells isolated from the peripheral blood of pregnant women

Immune cells from healthy individuals and ovarian cancer patients do not express endogenous MUC16 but instead bind the mucin that is present in the serum and the peritoneal fluid of ovarian cancer patients [11]. Higher levels of sMUC16 (measured as CA125) are also observed in the serum of pregnant women as compared to the healthy non-pregnant cohort [21, 25]. We therefore tested if similar to our observation with immune cells from ovarian cancer patients, sMUC16 also binds to immune cells of pregnant women.

MUC16 was immunoprecipitated from the lysates of PBMCs isolated at the time of delivery from healthy pregnant women. However, this mucin was not detected in the

immunoprecipitates from the lysates of healthy donor derived PBMCs (Fig. 1a). RT-PCR studies demonstrated that similar to ovarian cancer [12], the PBMCs from healthy pregnant women did not express endogenous MUC16 (Fig. 1b).

We have previously shown that sMUC16 binds to immune cells via Siglec-9 [11]. Binding of sMUC16 to Siglec-9 is rapid as shown by in vitro studies with Jurkat cells expressing this receptor (Fig. 2). The lack of MUC16 synthesis by immune cells of pregnant women suggested that similar to our observations in ovarian cancer, MUC16 is also captured on the surface of immune cells in pregnancy via Siglec-9. In subsequent experiments we therefore simultaneously monitored sMUC16 binding and Siglec-9 expression on the immune cells from pregnant women and preeclampsia patients by using multi-color flow cytometry.

sMUC16 binding pattern to immune cell subsets in pregnancy

All of our flow cytometry experiments require careful standardization of the instruments with calibration beads. We are therefore able to compare the binding pattern of sMUC16 and Siglec-9 expression in immune cells from pregnant subjects. Analysis of PBMC from healthy pregnant women indicated sMUC16 binding to B cells, NK cells and monocytes (Fig. 3). In the case of NK cells, equal levels of the mucin were observed on the $CD16^{pos}/CD56^{dim}$ and $CD16^{neg}/CD56^{br}$ subsets (Fig. 3). The sMUC16 binding pattern in pregnancy samples matched the expression of Siglec-9 on the two NK cell subsets, further suggesting that Siglec-9 is the predominant receptor for sMUC16 on these immune cells (Fig. 3).

Low level of sMUC16 was detected on $CD3^{pos}$ T cells that have previously been shown to express only low levels of Siglec-9 [17]. On the other hand, monocytes express high levels of Siglec-9 and were also strongly positive for sMUC16 (Fig. 3). In the case of B cells from pregnant women only 10% of the cells were double positive for sMUC16 and Siglec-9 even though approximately 80% of the $Siglec\text{-}9^{neg}$ B cells were positive for sMUC16 (Fig. 3). These data suggest that at least on the B cells, sMUC16 binding is Siglec-9-independent. Galectin-1 is another reported binding partner of MUC16 [26]. We have not found any correlation between Galectin-1 expression on B cells and sMUC16 (data not shown).

NK cell bound sMUC16 is detected even at low serum CA125 levels

Maximum serum CA125 levels are observed in the first trimester of pregnancy [22, 25]. We therefore monitored serum CA125 levels and immune cell bound sMUC16 in three women before and at 9 and 18 weeks of pregnancy. A gradual increase in serum CA125 was observed at 9 weeks of pregnancy [12]. Continued monitoring of these three patients at 18 weeks of the pregnancy (same gestation as that monitored in our previous study [12]) showed a trend toward lower serum CA125 levels (Fig. 4). However, levels of sMUC16 bound to immune cells continued to show an upward trend at 18 weeks of pregnancy even as the serum CA125 levels decreased (Fig. 4).

MUC16 binding to immune cells of patients with preeclampsia

Impairment of immune responses, especially those mediated by decidual NK cells, have been associated with the pathogenesis of preeclampsia [27, 28]. Given the demonstrated immunomodulatory roles of sMUC16 [7, 8], we next determined if there were any differences in the binding pattern of this mucin to immune cells in preeclamptic patients compared to healthy pregnant women. First, we confirmed the results of previous studies [22, 29, 30] that the serum CA125 levels were not significantly different between healthy pregnant women and preeclampsia patients (Fig. 5a). Similarly, there was no statistically significant difference in the percentage of monocytes (Fig. 5b) and NK cells (data not shown) in the PBMC of healthy pregnant women and preeclampsia patients.

Next we compared the binding of sMUC16 and expression of Siglec-9 on NK cells and monocytes from PBMC of preeclampsia patients and healthy pregnant subjects. Samples from healthy pregnant women described in Table 1 were analyzed in these experiments but did not include samples from patients who were already studied for experiments described in Fig. 3. The preeclampsia cohort included patients with both mild as well as severe disease, however a subgroup analysis was not performed in this initial study. Similar to our observations in healthy pregnant women, sMUC16 was also detected on NK cells and monocytes isolated from the PBMC of patients with preeclampsia (Fig. 6 and 7). The pattern of sMUC16 binding was distinct between the NK cell subsets of healthy pregnant women and preeclamptic women. Siglec-9 expression was comparable on the $CD16^{pos}/CD56^{dim}$ and $CD16^{neg}/CD56^{br}$ NK cells of the healthy pregnant women and the preeclampsia patients (Fig. 6a and b). However, a statistically significant increase in sMUC16 binding was observed on the $CD16^{pos}/CD56^{dim}$ and $CD16^{neg}/CD56^{br}$ NK cells of preeclampsia patients as compared to the healthy pregnant group (Fig. 6a and b). Significant differences were also observed in the case of monocytes isolated from healthy pregnant women and patients with preeclampsia. Increased sMUC16 binding was observed on monocytes of preeclampsia patients (Fig. 7). A trend for increased expression of Siglec-9 on the monocytes of preeclampsia patients was observed. However, the differences in expression levels of Siglec-9 between the two cohorts were not statistically significant ($p=0.649$; Fig. 7).

Discussion

Mucins are high molecular weight glycoproteins that exhibit important biological roles via their carbohydrate as well as protein epitopes [10, 31]. Probably the best characterized roles of the protein epitopes of mucins are the importance of the C-terminal domains of mucins such as MUC1 and MUC4 in intracellular signaling [32]. Oligosaccharides attached to the extracellular domains of mucins have also been shown to interact with different lectin receptors and influencing functions such as cell-cell adhesion and regulating immune responses [33-35]. MUC16 is, to date, the largest mucin identified in the human genome and similar to other mucins of its class plays a major biological role in normal as well as pathologic tissues via its extensive N-linked and O-linked oligosaccharide chains and its approximately 24,000 amino acid protein backbone [7, 10, 36-38].

Highlights of the biological activities of MUC16 reported at this point include its role in (a) facilitating ovarian tumor metastasis via its interactions with mesothelin, a GPI-anchored glycoprotein expressed on mesothelial cells and cancer cells [37-39], (b) allowing ovarian tumor cells to escape NK cell immune synapse formation and directly inhibiting the ability of NK cells to cytolyse cancer targets [7, 8], (c) imparting resistance to chemotherapy in ovarian cancer cells [40], (d) contributing to cell survival in cancer cells [41, 42], and (d) serving as a barrier to adhesion of trophoblasts to the endometrial epithelium and of bacteria to the corneal epithelia [43]. Considering this background, our observations that sMUC16 binds to immune cells during pregnancy and shows higher level of binding to the $CD16^{neg}/CD56^{br}$ NK cell subset and monocytes from preeclamptic patients are likely to have important physiologic consequences. First, sMUC16 binding to monocytes and $CD16^{pos}/CD56^{dim}$ NK cells is likely to inhibit the cytolytic activities of these immune cells, as shown in our previous study thereby ensuring suppression of maternal cytotoxic immune responses against the fetal tissues [8].

Second, the $CD16^{neg}/CD56^{br}$ NK cells exhibit low cytotoxicity and instead are high producer of cytokines [44]. This NK cell subset has been conclusively implicated as performing a pro-fetal growth function by expressing angiogenic cytokines that enhance trophoblastic invasion of the uterine spiral arterioles [28].

Recent data emerging from experiments with monocytic cell lines indicates that engagement of Siglec-9 induces these cells to increase expression of the immunosuppressive cytokine IL-10 [45]. This cytokine is known to promote placental angiogenesis. Thus sMUC16 binding to monocytes and NK cells may likely promote fetal development. Increased binding of sMUC16 to $CD16^{neg}/CD56^{br}$ NK cells and monocytes in preeclampsia may be a corrective maneuver to increase VEGF levels to promote angiogenesis in pregnancy.

Preeclampsia is a hypertensive disorder unique to pregnancy and affects 2-7% of pregnancies [23, 46]. It is a major cause of both maternal and fetal morbidity and mortality. Considered a leading cause of iatrogenic prematurity, preeclampsia is characterized by poor perfusion to multiple vital fetal organs which is reversed upon delivery. A primary pathological feature of preeclampsia is insufficient invasion of the spiral arterioles by the fetal trophoblast with subsequent widespread systemic endothelial dysfunction. $CD16^{neg}/CD56^{br}$ NK cells present in the decidua are implicated in successful trophoblastic invasion of the uterine spiral arterioles and the regulation of uterine blood flow [28].

Maternal hypertension is the primary mode of detection of preeclampsia. Increase in maternal blood pressure however occurs at a later stage in pregnancy. Attempts are therefore underway to identify biomarkers that can detect the onset of preeclampsia at an early stage [47-49]. Past reports have studied potential increases in the serum levels of CA125 and demonstrated no significant change in the levels of this marker between women with healthy pregnancies and those with preeclampsia [22, 29]. The studies presented here suggest a novel method for the identification of biomarkers that can predict preeclampsia. Based on this method we predict that binding of sMUC16 bound to circulating $CD16^{neg}/CD56^{br}$ NK cells and monocytes above a threshold value may serve as an important method for detection and monitoring of preeclampsia. Additionally, since sMUC16 also binds to $CD16^{pos}/CD56^{dim}$ NK cells and monocytes in ovarian cancer patients monitoring the levels of this mucin on immune cells may also prove important in detecting and monitoring ovarian cancer [11, 12]. Immune cell bound sMUC16 provides an advantage over the conventional serum CA125 assay as it does not primarily rely on the absolute concentration of the biomarker in the serum; a major obstacle since the concentration of the biomarker may vary based on its rapid turnover by the liver [50]; but instead relies on the measurement of the captured antigen on specific immune cell subsets.

Analysis of a large cohort of healthy donors, healthy pregnant women, and patients with ovarian cancer or preeclampsia will be necessary to establish threshold values for sMUC16 binding to NK cell subsets and monocytes to distinguish between healthy pregnant women and patients with preeclampsia. These studies are currently underway in our laboratories. Sequential samples obtained from ovarian cancer patients undergoing chemotherapy and pregnant diabetic women (a high risk group for preeclampsia) are being monitored for sMUC16 binding and Siglec-9 expression on immune cell subsets. These studies will be very useful in providing a profile of sMUC16 binding and Siglec-9 expression on immune cells and its potential in early identification and monitoring of ovarian cancer and preeclampsia.

Acknowledgments

VK-8 antibody used in these experiments was a generous gift from Dr. Beatrice Yin, Memorial Sloan Kettering. Funding for this research was provided by grants from, Congressionally Directed Medical Research Program of the Department of Defense (W81XWH-11-1-0181), UW School of Medicine and Public Health Wisconsin Partnership Program, The Robert Draper Technology Innovation Fund, a charitable donation from Jean McKenzie, Wisconsin Alliance for Ovarian Cancer, Department of Obstetrics and Gynecology, and National Institutes of Health (R21CA143616) grant to MSP and JPC. This investigation was supported by the National Institutes of Health under Ruth L. Kirschstein National Research Service Award 5T32HD041921 to C.T.T. and J.A.A.G. CTT is also the recipient of the Centennial Scholars Research grant and is BIRCWH scholar. Its contents are solely the

responsibility of the authors and do not necessarily represent the official views of the NIH. We are deeply grateful to Kathy Schell for her advice and help and acknowledge the support provided by the University of Wisconsin Comprehensive Cancer Centers Flow Cytometry facility which is supported by a core grant (CA14520) from the National Institutes of Health.

References

1. Bast RC Jr, Feeney M, Lazarus H, Nadler LM, Colvin RB, Knapp RC. Reactivity of a monoclonal antibody with human ovarian carcinoma. J Clin Invest. 1981; 68:1331–1337. [PubMed: 7028788]
2. Bast RC Jr, Klug TL, St John E, Jenison E, Niloff JM, Lazarus H, Berkowitz RS, Leavitt T, Griffiths CT, Parker L, Zurawski VR Jr, Knapp RC. A radioimmunoassay using a monoclonal antibody to monitor the course of epithelial ovarian cancer. N Engl J Med. 1983; 309:883–887. [PubMed: 6310399]
3. Rustin GJ. Use of CA-125 to assess response to new agents in ovarian cancer trials. J Clin Oncol. 2003; 21:187s–193s. [PubMed: 12743133]
4. Yin BW, Dnistrian A, Lloyd KO. Ovarian cancer antigen CA125 is encoded by the MUC16 mucin gene. Int J Cancer. 2002; 98:737–740. [PubMed: 11920644]
5. Yin BW, Lloyd KO. Molecular cloning of the ca125 ovarian cancer antigen. identification as a new mucin, muc16. J Biol Chem. 2001; 276:27371–27375. [PubMed: 11369781]
6. O'Brien TJ, Beard JB, Underwood LJ, Dennis RA, Santin AD, York L. The CA 125 gene: an extracellular superstructure dominated by repeat sequences. Tumour Biol. 2001; 22:348–366. [PubMed: 11786729]
7. Gubbels JA, Felder M, Horibata S, Belisle JA, Kapur A, Holden H, Petrie S, Migneault M, Rancourt C, Connor JP, Patankar MS. MUC16 provides immune protection by inhibiting synapse formation between NK and ovarian tumor cells. Mol Cancer. 2010; 9:11. [PubMed: 20089172]
8. Patankar MS, Yu J, Morrison JC, Belisle JA, Lattanzio FA, Deng Y, Wong NK, Morris HR, Dell A, Clark GF. Potent suppression of natural killer cell response mediated by the ovarian tumor marker CA125. Gynecol Oncol. 2005; 99:704–713. [PubMed: 16126266]
9. O'Brien TJ, Beard JB, Underwood LJ, Shigemasa K. The CA 125 gene: a newly discovered extension of the glycosylated N- terminal domain doubles the size of this extracellular superstructure. Tumour Biol. 2002; 23:154–169. [PubMed: 12218296]
10. Hattrup CL, Gendler SJ. Structure and Function of the Cell Surface (Tethered) Mucins. Annu Rev Physiol. 2007; 70:431–457. [PubMed: 17850209]
11. Belisle JA, Horibata S, Gubbels JA, Petrie S, Kapur A, Andre S, Gabius HJ, Rancourt C, Connor J, Paulson JC, Patankar MS. Identification of Siglec-9 as the receptor for MUC16 on human NK cells, B cells, and monocytes. Mol Cancer. 2010; 9:118. [PubMed: 20497550]
12. Belisle JA, Gubbels JA, Raphael CA, Migneault M, Rancourt C, Connor JP, Patankar MS. Peritoneal natural killer cells from epithelial ovarian cancer patients show an altered phenotype and bind to the tumour marker MUC16 (CA125). Immunology. 2007; 122:418–429. [PubMed: 17617155]
13. Angata T, Varki A. Cloning, characterization and phylogenetic analysis of Siglec-9, a new member of the CD33-related group of Siglecs. Evidence for co-evolution with sialic acid synthesis pathways. J Biol Chem. 2000; 275:22127–22135. [PubMed: 10801860]
14. Yamaji T, Teranishi T, Alphey MS, Crocker PR, Hashimoto Y. A small region of the natural killer cell receptor, Siglec-7, is responsible for its preferred binding to alpha2,8-disialyl and branched alpha2,6-sialyl residues: a comparison with Siglec-9. J Biol Chem. 2001; 10:6324–6332. [PubMed: 11741958]
15. Zhang JQ, Nicoll G, Jones C, Crocker PR. Siglec-9, a novel sialic acid binding member of the immunoglobulin superfamily expressed broadly on human blood leukocytes. J Biol Chem. 2000; 275:22121–22126. [PubMed: 10801862]
16. Avril T, Floyd H, Lopez F, Vivier E, Crocker PR. The membrane-proximal immunoreceptor tyrosine-based inhibitory motif is critical for the inhibitory signaling mediated by Siglecs-7 and -9, CD33-related Siglecs expressed on human monocytes and NK cells. J Immunol. 2004; 173:6841–6849. [PubMed: 15557178]

17. Ikehara Y, Ikehara SK, Paulson JC. Negative regulation of T cell receptor signaling by Siglec-7 (p70/AIRM) and Siglec-9. J Biol Chem. 2004; 279:43117–43125. [PubMed: 15292262]
18. O'Brien TJ, Hardin JW, Bannon GA, Norris JS, Quirk JG Jr. CA 125 antigen in human amniotic fluid and fetal membranes. Am J Obstet Gynecol. 1986; 155:50–55. [PubMed: 3460340]
19. Quirk JG Jr, Brunson GL, Long CA, Bannon GA, Sanders MM, O'Brien TJ. CA 125 in tissues and amniotic fluid during pregnancy. Am J Obstet Gynecol. 1988; 159:644–649. [PubMed: 3273362]
20. Hardardottir H, Parmley TH 2nd, Quirk JG Jr, Sanders MM, Miller FC, O'Brien TJ. Distribution of CA 125 in embryonic tissues and adult derivatives of the fetal periderm. Am J Obstet Gynecol. 1990; 163:1925–1931. [PubMed: 2256504]
21. Niloff JM, Knapp RC, Schaetzl E, Reynolds C, Bast RC Jr. CA125 antigen levels in obstetric and gynecologic patients. Obstet Gynecol. 1984; 64:703–707. [PubMed: 6208522]
22. Bon GG, Kenemans P, Verstraeten AA, Go S, Philipi PA, van Kamp GJ, van Geijn HP, van Vugt JM. Maternal serum Ca125 and Ca15-3 antigen levels in normal and pathological pregnancy. Fetal Diagn Ther. 2001; 16:166–172. [PubMed: 11316933]
23. Sibai B, Dekker G, Kupferminc M. Pre-eclampsia. Lancet. 2005; 365:785–799. [PubMed: 15733721]
24. Goodell CAR, Belisle JA, Gubbels JAA, Migneault M, Rancourt C, Connor JP, Kunnimalaiyaan M, Kravitz R, Tucker W, Zwick M, Patankar MS. Characterization of the Tumor Marker Muc16 (CA125) Expressed by Murine Ovarian Tumor Cell Lines and Identification of a Panel of Cross-Reactive Monoclonal Antibodies. Journal of Ovarian Research. 2009; 2:8. [PubMed: 19538730]
25. Seki K, Kikuchi Y, Uesato T, Kato K. Increased serum CA 125 levels during the first trimester of pregnancy. Acta Obstet Gynecol Scand. 1986; 65:583–585. [PubMed: 3467532]
26. Seelenmeyer C, Wegehingel S, Lechner J, Nickel W. The cancer antigen CA125 represents a novel counter receptor for galectin-1. J Cell Sci. 2003; 116:1305–1318. [PubMed: 12615972]
27. Ahn H, Park J, Gilman-Sachs A, Kwak-Kim J. Immunologic Characteristics of Preeclampsia, a Comprehensive Review. Am J Reprod Immunol. 2011; 65:377–394. [PubMed: 20825381]
28. Hanna J, Goldman-Wohl D, Hamani Y, Avraham I, Greenfield C, Natanson-Yaron S, Prus D, Cohen-Daniel L, Arnon TI, Manaster I, Gazit R, Yutkin V, Benharroch D, Porgador A, Keshet E, Yagel S, Mandelboim O. Decidual NK cells regulate key developmental processes at the human fetal-maternal interface. Nat Med. 2006; 12:1065–1074. [PubMed: 16892062]
29. Schrocksnadel H, Daxenbichler G, Artner E, Steckel-Berger G, Dapunt O. Tumor markers in hypertensive disorders of pregnancy. Gynecol Obstet Invest. 1993; 35:204–208. [PubMed: 7687229]
30. de Groot CJ, O'Brien TJ, Taylor RN. Biochemical evidence of impaired trophoblastic invasion of decidual stroma in women destined to have preeclampsia. Am J Obstet Gynecol. 1996; 175:24–29. [PubMed: 8694068]
31. Hollingsworth MA, Swanson BJ. Mucins in cancer: protection and control of the cell surface. Nat Rev Cancer. 2004; 4:45–60. [PubMed: 14681689]
32. Bafna S, Kaur S, Batra SK. Membrane-bound mucins: the mechanistic basis for alterations in the growth and survival of cancer cells. Oncogene. 2010; 29:2893–2904. [PubMed: 20348949]
33. Swanson BJ, McDermott KM, Singh PK, Eggers JP, Crocker PR, Hollingsworth MA. MUC1 is a counter-receptor for myelin-associated glycoprotein (Siglec-4a) and their interaction contributes to adhesion in pancreatic cancer perineural invasion. Cancer Res. 2007; 67:10222–10229. [PubMed: 17974963]
34. Iida S, Yamamoto K, Irimura T. Interaction of human macrophage C-type lectin with O-linked N-acetylgalactosamine residues on mucin glycopeptides. J Biol Chem. 1999; 274:10697–10705. [PubMed: 10196140]
35. Patankar M, Gubbels JAA, Felder M, Connor JP. Immunomodulatory properties of glycoproteins in ovarian cancer. Frontiers in Biosciences (In press). 2012
36. Blalock TD, Spurr-Michaud SJ, Tisdale AS, Heimer SR, Gilmore MS, Ramesh V, Gipson IK. Functions of MUC16 in corneal epithelial cells. Invest Ophthalmol Vis Sci. 2007; 48:4509–4518. [PubMed: 17898272]
37. Gubbels JA, Belisle J, Onda M, Rancourt C, Migneault M, Ho M, Bera TK, Connor J, Sathyanarayana BK, Lee B, Pastan I, Patankar MS. Mesothelin-MUC16 binding is a high affinity,

N-glycan dependent interaction that facilitates peritoneal metastasis of ovarian tumors. Mol Cancer. 2006; 5:50. [PubMed: 17067392]

38. Rump A, Morikawa Y, Tanaka M, Minami S, Umesaki N, Takeuchi M, Miyajima A. Binding of ovarian cancer antigen CA125/MUC16 to mesothelin mediates cell adhesion. J Biol Chem. 2004; 279:9190–9198. [PubMed: 14676194]
39. Chang K, Pastan I. Molecular cloning of mesothelin, a differentiation antigen present on mesothelium, mesotheliomas, and ovarian cancers. Proc Natl Acad Sci U S A. 1996; 93:136–140. [PubMed: 8552591]
40. Boivin M, Lane D, Piche A, Rancourt C. CA125 (MUC16) tumor antigen selectively modulates the sensitivity of ovarian cancer cells to genotoxic drug-induced apoptosis. Gynecol Oncol. 2009; 115:407–413. [PubMed: 19747716]
41. Reinartz S, Failer S, Schuell T, Wagner U. CA125 (MUC16) gene silencing suppresses growth properties of ovarian and breast cancer cells. Eur J Cancer. 2011 epub ahead of print.
42. Lakshmanan I, Ponnusamy MP, Das S, Chakraborty S, Haridas D, Mukhopadhyay P, Lele SM, Batra SK. MUC16 induced rapid G2/M transition via interactions with JAK2 for increased proliferation and anti-apoptosis in breast cancer cells. Oncogene.
43. Gipson IK, Blalock T, Tisdale A, Spurr-Michaud S, Allcorn S, Stavreus-Evers A, Gemzell K. MUC16 is Lost from the Uterodome (Pinopode) Surface of the Receptive Human Endometrium: In Vitro Evidence That MUC16 Is a Barrier to Trophoblast Adherence. Biol Reprod. 2007; 78:134–142. [PubMed: 17942799]
44. Farag SS, Caligiuri MA. Human natural killer cell development and biology. Blood Rev. 2006; 20:123–137. [PubMed: 16364519]
45. Ando M, Tu W, Nishijima K, Iijima S. Siglec-9 enhances IL-10 production in macrophages via tyrosine-based motifs. Biochem Biophys Res Commun. 2008; 369:878–883. [PubMed: 18325328]
46. Sankaralingam S, Arenas IA, Lalu MM, Davidge ST. Preeclampsia: current understanding of the molecular basis of vascular dysfunction. Expert Rev Mol Med. 2006; 8:1–20. [PubMed: 16438753]
47. Kuc S, Wortelboer EJ, van Rijn BB, Franx A, Visser GH, Schielen PC. Evaluation of 7 serum biomarkers and uterine artery Doppler ultrasound for first-trimester prediction of preeclampsia: a systematic review. Obstet Gynecol Surv. 66:225–239. [PubMed: 21756405]
48. Cuckle HS. Screening for pre-eclampsia--lessons from aneuploidy screening. Placenta. 32(Suppl):S42–48. [PubMed: 21257082]
49. Romero R, Kusanovic JP, Than NG, Erez O, Gotsch F, Espinoza J, Edwin S, Chefetz I, Gomez R, Nien JK, Sammar M, Pineles B, Hassan SS, Meiri H, Tal Y, Kuhnreich I, Papp Z, Cuckle HS. First-trimester maternal serum PP13 in the risk assessment for preeclampsia. Am J Obstet Gynecol. 2008; 199:122 e121–122 e111. [PubMed: 18539259]
50. Wahrenbrock MG, Varki A. Multiple hepatic receptors cooperate to eliminate secretory mucins aberrantly entering the bloodstream: are circulating cancer mucins the “tip of the iceberg”? Cancer Res. 2006; 66:2433–2441. [PubMed: 16489050]
51. Lloyd KO, Yin BW, Kudryashov V. Isolation and characterization of ovarian cancer antigen CA 125 using a new monoclonal antibody (VK-8): identification as a mucin-type molecule. Int J Cancer. 1997; 71:842–850. [PubMed: 9180155]

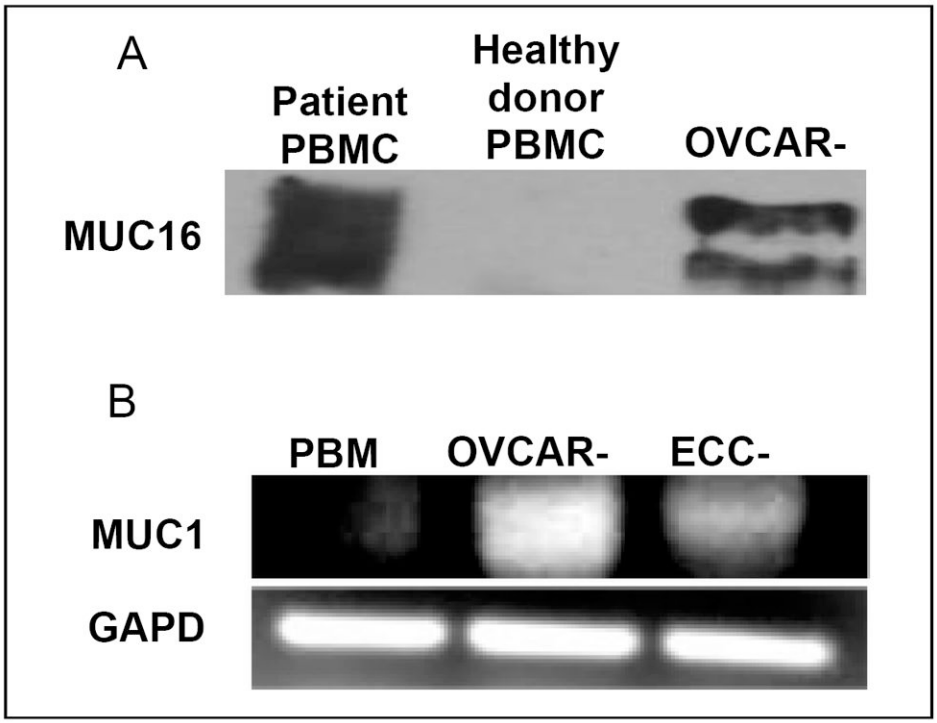


Fig. 1.

sMUC16 is captured but not expressed by immune cells of pregnant women. a, PBMC isolated from healthy pregnant women (lane 1), and a non-pregnant individual were lysed and immunoprecipitated with the anti-MUC16 antibody, VK8 [51]. The immunoprecipitated material (5 μg/lane) was analyzed by western blotting and VK8 was used for detection of the mucin. Lysate of ovarian cancer cell line OVCAR-3 was used as positive control (lane 3). b, mRNA was isolated from PBMCs from women with healthy pregnancy and MUC16 was detected by qPCR. mRNA from OVCAR-3 and ECC-1 cells were used as positive controls.

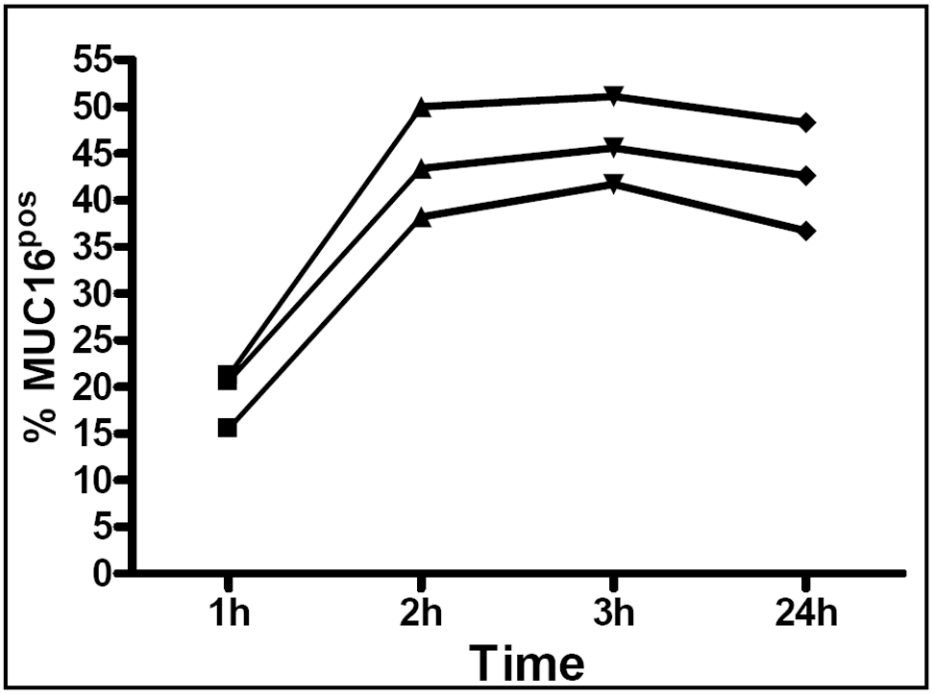


Fig. 2.

Rapid binding of sMUC16 to immune cells via Siglec-9. Jurkat cells transfected with Siglec-9 were incubated with sMUC16 partially purified from the conditioned media of OVCAR-3 cells. Binding of the mucin to the Siglec-9 expressing Jurkat cells was determined by flow cytometry.

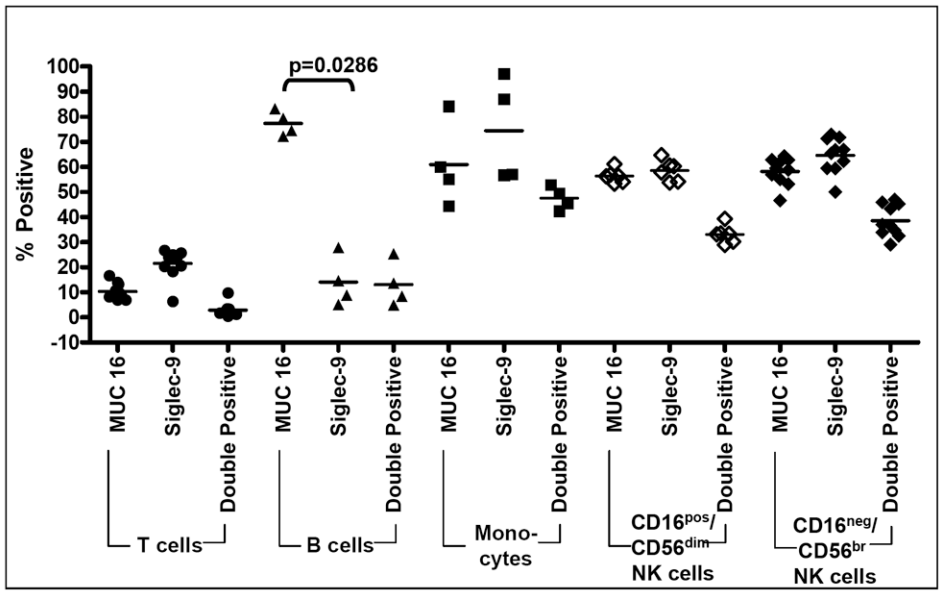
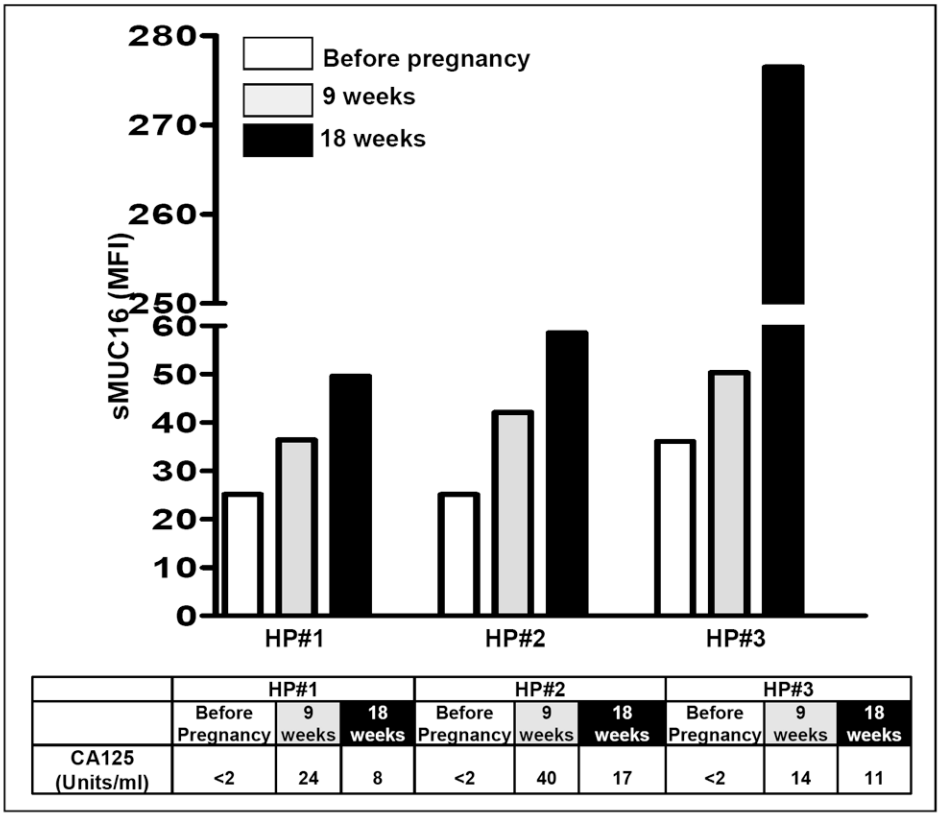


Fig. 3.

sMUC16 predominantly binds to Siglec-9 expressing cells isolated from the peripheral blood of healthy pregnant women. sMUC16 binding to T cells, B cells, NK cells, and monocytes from PBMC of healthy pregnant women was determined by flow cytometry. Siglec-9 positive events were gated and the binding of sMUC16 to these immune cell subsets was determined. Immune cells from each patient were analyzed by flow cytometry in duplicates and the data for each reading is shown.



	HP#1			HP#2			HP#3		
	Before Pregnancy	9 weeks	18 weeks	Before Pregnancy	9 weeks	18 weeks	Before Pregnancy	9 weeks	18 weeks
CA125 (Units/ml)	<2	24	8	<2	40	17	<2	14	11

Fig. 4.

Increasing levels of immune cell bound sMUC16 during pregnancy. Three healthy pregnant women (HP#1, HP#2, and HP#3) were analyzed for serum CA125 levels (Units/ml shown below bar graph) and sMUC16 bound to peripheral blood immune cells (bar graph) isolated at weeks 9 and 18 of pregnancy. Immune cells from these three donors prior to their pregnancy were also available and were also tested for bound sMUC16. Bar graphs show the mean fluorescence intensity (MFI) for the total sMUC16 positive events in each sample used in this analysis.

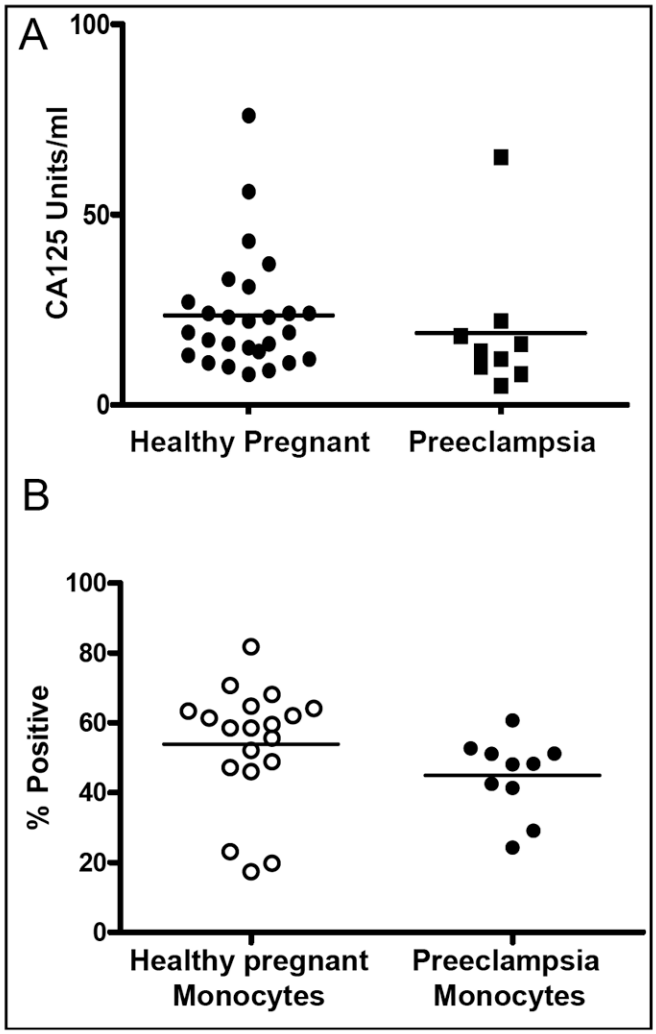


Fig. 5.

Healthy pregnant women and preeclampsia patients have comparable levels of serum CA125 and monocytes. A, Serum samples were drawn from healthy pregnant (n=27) and preeclampsia patients (n=9) and CA125 levels in each sample were determined using a clinical assay. b, Healthy pregnant women and preeclampsia patients have comparable levels of peripheral blood monocytes. Comparable levels of immune cells were present in the peripheral blood of healthy pregnant women and preeclampsia patients. Representative data for peripheral blood monocytes is shown here. Data shows the total number of $CD3^{neg}$/$CD16^{pos}$/$CD14^{pos}$ monocyte events as determined by flow cytometry in each blood sample.

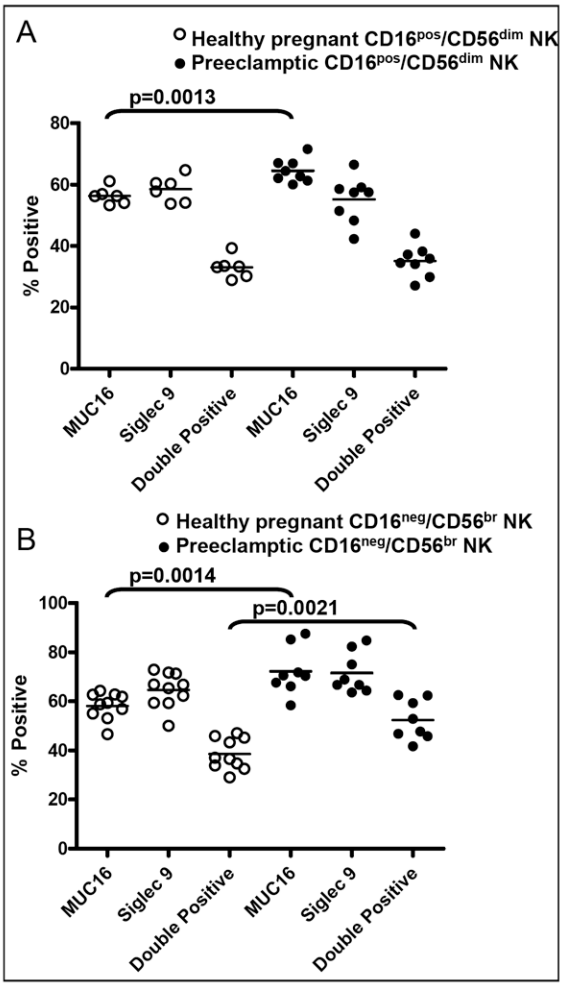


Fig. 6.

Differential binding to sMUC16 to NK cell subsets of preeclaptic patients. Peripheral blood mononuclear cells isolated from healthy pregnant women and preeclampsia patients at the time of delivery were subjected to multi-color flow cytometry. The $CD16^{pos}/CD56^{dim}$ and $CD16^{neg}/CD56br$ NK cell subsets were identified as described in our previous work [11, 12]. sMUC16 was detected using the VK-8 antibody and percentage of sMUC16 and Siglec-9 positive cells were identified. The flow cytometer was calibrated using fluorescent bead standards allowing us to compare the data obtained from all samples from experiments conducted on different days. Duplicate readings for each patient are shown.

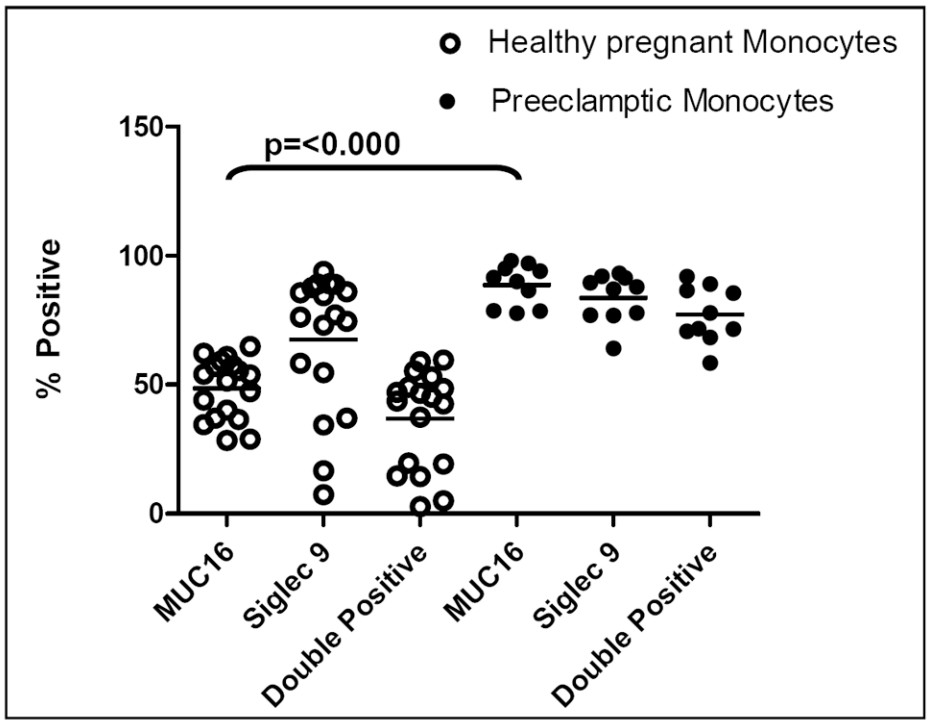


Fig. 7.

Differential binding to sMUC16 to monocytes of preeclamptic patients. Peripheral blood mononuclear cells isolated from healthy pregnant women and preeclampsia patients at the time of delivery were subjected to multi-color flow cytometry. Monocytes were identified as described [11, 12] and percent of sMUC16 (detected using VK-8 antibody) and Siglec-9 positive events were identified. The flow cytometer was calibrated using fluorescent bead standards allowing us to compare the data obtained from all samples from experiments conducted on different days. Duplicate readings for each patient are shown.

Table 1

Characteristics	Controls (n=17)	Preeclamptics (n=9)	pValue
Age	33.45 (22-42)	27.4 (19-40)	*0.0095
Body Mass Index (BMI)	32.45 (26-47)	37.2 (27-51)	0.0555
Gestational Age at Delivery (GA)	38.75 (37.2-41)	35.89 (28.1-38.3)	*0.0009

1. Introduction

With the advent of high-throughput mass spectrometry (MS), shotgun proteomics has been employed as a major tool to analyze biological samples and produce thousands of MS or tandem MS spectra [1–3]. The method of choice for annotating these spectra with sequences are currently database search engines such as Mascot, SEQUEST, OMSSA and X!Tandem [4–7]. The algorithms generally score the similarities between the experimental and theoretical spectra and rank the best match with the highest score as the predicted peptide spectrum match (PSM). However, the top PSM is not necessarily correct due to the scoring scheme and quality of a spectrum. Target/decoy search strategy and the resulting false discovery rate (FDR) calculation is used to assess the confidence of reported PSMs[8]. However, there is a tradeoff between sensitivity and accuracy of peptide or protein identifications that FDR has to manage [9, 10]. In order to increase sensitivity while maintain accuracy one can incorporate retention time predictors [11–22] as a post-Mascot analysis tool to increase confidence for peptide identifications.

The retention time (RT) of a peptide is defined as the elapsed time between the time of injection and the time of elution of the peak maximum. Previous studies demonstrated that the retention time of a peptide is the function of various peptide parameters, including amino acid composition [23], N-terminal or C-terminal residues[14], location of amino acids within the primary structure [16], peptide length or mass[12], and hydrophobicity [20]. Many sophisticated models have been constructed to predict retention time and used predicted RT to improve peptide identification. For example, Krokhin et al. [14] trained a linear model (SSRC) by linearly correlating RT with a comprehensive hydrophobicity which integrates residue's hydrophobicity and structural and positional effects. Strittmatter et al. [21] proposed an artificial neural network (ANN) peptide RT prediction model by using positional amino acid information to yield a 16% increase in peptide identification for a complex sample (human plasma) [16]. Klammer et al.[22] adopted a support vector regressor dynamically trained for each chromatographic run, with which 50% more positive peptide identifications were obtained at a false positive rate of 3%. Although a great deal of effort has been made in this field to improve protein identifications in shotgun proteomics, there are still some challenging issues to be addressed. The comparison of the previously published models indicates that the static linear model depends on the chromatographic condition and thus prediction bias would occur when the model is used for a different condition. ANN model needs an extremely large dataset (~345000 training examples) [21] that is often impractical for application. Dynamic SVR model is suitable for relatively small dataset and avoids the RT variation between different chromatographic runs. However, its performance is modest compared to the other two models. In addition, the deviation between predicted and experimental RTs (Δ RT) is favorably used to filter out false positives when applying the trained RT predictor to real data. However, there are limitations in previous approaches to determine a suitable Δ RT threshold. With SSRC [24], the Δ RT threshold was determined by tentatively checking recovery of peptide predictions with varying arbitrary Δ RT values like ±4, ±2, ±1 min. In [22], optimal Δ RT threshold was selected from a range of Δ RT values (0–240 min) at which the highest numbers of true positives were obtained across the largest number of FDR values (in a range of 0.5–10%). These approaches could lead to under or overestimation of peptide identifications by unsuitable Δ RT threshold. Hence, it is necessary to develop a state-of-the-art RT predictor which can increase the sensitivity to maximize the number of predictions while ensure the accuracy of peptide identifications at the same time.

Given that a dynamic SVR model is more universal and practical for real application, we developed a SVR based RT predictor (RT-SVR) to be used in conjunction with Mascot search results obtained from 2D LC-MS/MS experiments. Our proposed RT-SVR model

was constructed with multiple peptide spectral matches (PSM) which were obtained from Mascot search results (at FDR~1%) for each run. When applying the trained RT-SVR model to real data for examining peptide identifications, instead of choosing a Δ RT threshold arbitrarily or trained with a set of FDR, we introduced a method called q value assessment to define a dynamic Δ RT threshold that improves the confidence of evaluation for peptide identifications. By using this statistical method, q value rather than Δ RT is employed as the cut-off criteria to filter out the false positives. q value metric was first proposed by Storey et al.[25] to analyze genomic data and later on it was revised and applied to MS-based proteomics by Kail et al.[26, 27]. The q value can be understood as the minimal FDR at which a peptide spectra match (PSM) can be accepted. In practice, q value can be associated with any PSM in a dataset. Since q value is calculated from all PSMs in a dataset, it is considered as a statistical result for the whole dataset like FDR. Previous studies have shown that q value is equivalent to FDR estimation and no bias will be introduced toward under or overestimation [28, 29]. Thus, q value assessment is viewed as a more accurate and reliable estimation of error rate. In our study, a modified q value was calculated based on target and decoy PSMs and then was assigned to a peptide prediction. Finally, we can unambiguously filter out the false positives for a given q value threshold such as 0.01 (Figure 1). By applying our strategy to proteomic analysis of the natural killer leukemia cell line (NKL), the trained RT-SVR models for all datasets obtained from sample fractions show high performance with R value above 0.900. The peptide and protein identifications increase by up to 89.4% and 83.5% respectively in comparison with Mascot search results (at FDR 1%) with a q value of 0.01. Our results thus demonstrate the utility of the RT-SVR with q value assessment as a robust and reliable method for post-Mascot analysis in proteomic applications.

In addition, we combined RT-SVR and Mascot score screening (Mascot Identity Threshold) to rescue those peptide identifications missed by RT-SVR. This combined RT-SVR method yields more peptide and protein identifications.

In order to evaluate the general applicability of our RT-SVR strategy we applied the model to an independent set of large-scale yeast proteomic data acquired using a Thermo LTQ mass spectrometer (downloaded from PeptideAtlas (PAe001337) and processed by the combined RT-SVR). As a comparison, 566 unique proteins were predicted at a q value of 0.01 in contrast with 470 with Mascot (MIH, FDR 1%) and 499 unique proteins reported by Trans-Proteomic Pipeline (TPP, probability filter 0.010). This result suggests that the RT-SVR model is independent of instruments used for shotgun proteomics and is generally applicable to proteomic data analysis acquired on multiple mass spectrometric platforms.

RT-SVR was written in Java. The windows-based graphical user interface can be freely downloaded from http://pages.cs.wisc.edu/~yadi/bioinfo/rtsvr/rtsvr.html.

2. Materials and methods

2.1. Sample preparation for proteomic analysis

10^7 NKL cells were harvested and washed three times with ice-cold PBS. Cells were lysed with 100μl RIPA lysis buffer (Formulation: 50 mM Tris-HCl (pH7.4), 150 mM NaCl, 0.1% SDS, 1% NP-40) on ice for 20 minutes with 20 seconds of sonication at the beginning. Cellular debris was removed by centrifugation for 30 min at 16,100 ×g at 4°C. Supernatants were collected and protein concentrations were measured using a BCA protein assay kit (Pierce). 50μg of protein was used for acetone precipitation. Acetone (precooled to −80°C) was added gradually (with intermittent vortexing) to the protein extract to a final concentration of 80% (v/v). The solution was then incubated at −20 °C for 60 minutes and centrifuged at 16,100 ×g for 15 minutes. The supernatant was decanted, and the pellet was

carefully washed twice using cold acetone to ensure the efficient removal of detergent. The residual acetone was evaporated at ambient temperature. The pellet was dissolved and denatured with 8M urea in 25 mM ammonium bicarbonate buffer, and reduced by incubating with 50 mM DTT at 37°C for 1 hour. The reduced proteins were alkylated for 1 hour in darkness with 100 mM iodoacetamide. The alkylation reaction was quenched by adding DTT to a final concentration of 50 mM. The samples were diluted to a final concentration of 1 M urea. Trypsin was added to the sample at a 30:1 protein to trypsin mass ratio. The sample was incubated at 37°C overnight.

2.2. Off-line first dimension HPLC

Tryptic digests were injected onto Waters Alliance HPLC (Waters) with a high pH-stable RP column (Phenomenex Gemini C18, 150 × 2.1mm, 3 micron) at a flow rate of 150μL/min. The peptides were eluted with a gradient from 5 to 45% solvent B over 60 minutes (Solvent A: 100mM ammonium formate, pH 10; Solvent B: acetonitrile (ACN)). Fractions were collected every 3 min for 60 min. Collected fractions were dried by Speedvac and reconstituted in 30 μL of 0.1% formic acid. 5 μL of each of the 20 fractions were subjected to nanoLC-MS/MS.

2.3. LC-ESI ion trap mass spectrometry and MS/MS analysis

20 fractions collected from high pH RPLC were analyzed using amaZon ion trap mass spectrometer (Bruker Daltonics, Germany) equipped with Eksigent nanoLC-Ultra system (Dublin, CA). For the chromatographic separation, solvent A consisted of 0.1% formic acid in water and solvent B consisted of 0.1% formic acid in ACN. 5 μL of each fraction is injected onto an Agilent Technologies Zorbax 300 SB-C18 5 μm, 5×0.3 mm trap cartridge (Santa Clara, CA) at a flow rate of 5 μL/min for 5 minutes at 95% A 5% B, followed by peptides separation performed on Waters 3μm Atlantis dC18 75 μm × 150 mm analytical column (Milford, MA) using gradient from 0 to 45% solvent B at 250 nL/min over 90 minutes. Acquisition of precursor ions and MS/MS spectra was performed using the parameters as indicated below:

Smart parameter setting (SPS) was set to 700 *m/z*, compound stability and trap drive level were set at 100%. Dry gas temperature, 125°C, dry gas, 4.0 L/min, capillary voltage, −1300 V, end plate offset, −500V, MS/MS fragmentation amplitude, 1.0V, and Smart Fragmentation set at 30–300%. Data were generated in data dependent mode with strict active exclusion set after two spectra and released after one minute. MS/MS spectra were obtained via collision induced dissociation (CID) fragmentation for the six most abundant MS ions. For MS generation the ICC target was set to 200,000, maximum accumulation time, 50.00 ms, one spectrometric average, rolling average, 2, acquisition range of 300–1500 *m/z*, and scan speed (enhanced resolution) of 8,100 *m/z* s^{-1}. For MS/MS generation the ICC target was set to 300,000, maximum accumulation time, 50.00 ms, two spectrometric averages, acquisition range of 100–2000 *m/z*, and scan speed (Ultrascan) of 32,000 *m/z* per second.

2.4. MS/MS database search

MS/MS spectra were converted into mgf formatted files by DataAnalysis (Ver 4.0, Bruker Daltonics Bremen, Germany). Deviations in parameters from the default Protein Analysis in DataAnalysis were as follows: intensity threshold, 1000, maximum number of compounds, 1E9, and retention time window 0.001 minute. The resulting mgf files were then searched against the Human SwissProt database (SwissProt_57.5.fasta) with Mascot 2.2.06. The parameters and conditions were set as the following: tryptic digestion, maximum 3 missed cleavages, carbamidomethylation of cysteine as the fixed modification, oxidation of methionine as the variable modification, peptide mass tolerance of 100 ppm, fragment mass

tolerance of 0.6 Da, +1, +2 and +3 chosen for charge state. In this study, a simultaneous target-decoy search strategy (automatic decoy search) was adopted for FDR estimation. During the search, every time a protein sequence from the target database is tested, a random sequence of the same length is automatically generated and tested. Set “a bold red peptide required” for protein assembly.

2.5. Extracting training and test datasets

We used Mascot results obtained at a FDR of 1% as the source of training and test datasets. The FDR can be calculated as the number of decoy matches divided by the number of target matches. We accepted the peptide-spectra matches (PSM) above identity threshold as confident identifications when FDR is equal or less than 1%. By adjusting the significance level, p value defined by Mascot, the FDR can be controlled at ~1%. Following this, the resulting peptide identifications were exported as csv formatted files. A custom-written java script was then used to process the exported results and extract those identifications with scores above identity threshold. The corresponding retention times were also included. Finally, the set of peptides and associated retention times were randomly split according to a ratio of 3:1 to form the training and test datasets for each chromatographic run. No data was allowed to be included in both training and test datasets to avoid overfitting. The fraction with 100 or less PSMs at a FDR of 1% was not used for this study because the performance of a support vector regressor (SVR) would be deteriorated greatly if it was generated by a small number of training examples [30]. Via mascot search, there is no peptide contained in the first 1/4 and last 1/4 of the 20 fractions. We used the middle 10 fractions for our study. After checking the number of peptide predictions in each of 10 fractions, we found that 9 fractions met the criteria (confident PSMs >100) and they were used as the data source for dataset #1–9. In addition, we selected the last fraction as source for dataset #10 (containing 77 PSMs) to test if the performance of a SVR trained with small size of data could be deteriorated.

2.6. Constructing dynamic support vector regressor and performance analysis

Given the variations of retention times for a specific analyte with different chromatographic runs, a dynamic retention time regressor is needed to eliminate this bias for each chromatographic run [22]. A support vector regressor learns a function that relates a dependant variable, here retention time, to a set of independent variables (we called features). To generate the independent features, each data point including peptide and the associated retention time in training dataset was rewritten as 45-element support vectors. The 45 elements were treated as features composed of the following: 20 features representing the total number of each amino acid residue in the peptide sequence; 20 binary features representing the identity of the N-terminal amino acid residue; 2 binary features standing for the identity of C-termini (R or K, both 0 if no R/K); one feature for peptide mass; one feature for peptide length and the class feature for retention time. We used the same method to rewrite the data points in test dataset.

A 10-fold cross validation was set to optimize the hyper-parameter selection. Two kernel functions were used: a linear kernel which can report the weight of each feature and a RBF kernel (Gaussian function) which can produce a more flexible and successful regressor.

The performance of SVR model was evaluated by Pearson’s correlation, which measures the correlation coefficient, R value between the predicted and experimental retention times. The R value for two datasets X and Y of length n is given by the following formula,

$$R=\frac{Cov(X,Y)}{\sigma_X\sigma_Y}$$

Where

$$Cov(X,Y)=n\sum x_iy_i-\sum x_i\sum y_i$$
$$\sigma_X=\sqrt{n\sum x_i^2-(\sum x_i)^2}\qquad \sigma_Y=\sqrt{n\sum y_i^2-(\sum y_i)^2}$$

The whole procedure, including training and test dataset extracting, model learning and performance analysis, was repeated 5 times to eliminate the data variance[31].

2.7. Application of RT-SVR model and statistical analysis for peptide predictions

For each of chromatographic run #1 to #9, we adjusted p value to 0.10 when employing Mascot to do a simultaneous target-decoy search (automatic decoy search). The resulting Mascot report including target and decoy peptide sequence matches (PSM) was exported as .csv formatted files. Following by this, experimental retention time for each PSM was picked up and put in a new column “RT/min” in .csv files. The revised .csv files were then processed by the corresponding RT-SVR model during which the theoretical RTs were predicted and RT errors (ΔRT, differences between experimental and theoretical RT’s) were calculated. The RT error was used to calculate q value which was used to statistically assess the accuracy of peptide predictions.

The following procedure describes the calculation of q value.

Denote the RT errors (Δ RT) of target PSMs f_1, f_2, …, f_{mf} and the RT errors of decoy PSMs d_1, d_2, …, d_{md}. For a given Δ RT threshold *t*, the false discovery rate (FDR) can be estimated as

$$FDR(t)=\frac{E(FP(t))}{E(P(t))}$$

E(FP(t)) is the expected value of the number of false positives and E(P(t)) is the expected value of the number of positives. E(FP(t)) and E(P(t)) can be experimentally determined by doing a simultaneous target-decoy search. E(FP(t)) is the number of experimentally accepted decoy PSMs, which is denoted as $E(FP(t))=|\{d_i \le t; i=1,2,\ldots, m_d\}|$. E(P(t)) is the number of experimentally accepted target PSMs denoted as $E(P(t))=|\{f_i \le t, i=1,2\ldots, m_f\}|$. Therefore, we can rewrite the formula of FDR estimation as

$$FDR(t)=\frac{|\{d_i \le t, i=1,2,\ldots,m_d\}|}{|\{f_i \le t, i=1,2,\ldots,m_f\}|}$$

For a given target PSM with a RT error of *s*, the associated q value is defined as the minimal FDR value as shown in the following equation:

$$q(s)=\min_{t\geq s}\{FDR(t)\}$$

To calculate q value for a PSM with a RT error of *s*, 1) compare all other RT errors to *s*; 2) compute FDR by setting the RT error as threshold if a RT error ≥ *s*. 3) compare all calculated FDRs and choose the smallest one as the q value for the PSM with a RT error of *s*.

Throughout the paper we calculate q values at the PSM level; i.e., the same peptide can be reported as a target or decoy identification multiple times. Through the definition, we can consider q value as the minimal FDR at which the PSM can be accepted.

3. Results and Discussion

3.1. Support vector regressor and performance analysis

The information about training and test datasets used for this study is summarized in Table 1 in which the number of confident peptide identifications is listed for each chromatographic run. The size ratio of training dataset to test data set is randomly within 3:1 to 4:1.

We used Alex Smola and Bernhard Scholkopf's sequential minimal optimization algorithm to train a support vector regression model (SMOreg). Two kernel functions were utilized in this study so that we could compare their performance and then choose the better one for application to protein identifications. Figure 2 demonstrates the comparison of the performance of both kernels. The Bland-Altman plots[32] were made with the results from dataset #1. From the graph, most deviations in retention time (RT error) fall into the region of 95% confidence limit of bias (mean±2SD). The mean and the region of mean±2SD (0.07±1.76 min) with Gaussian kernel are smaller than those (0.33+2.60 min) with linear kernel. Moreover, the R value with Gaussian Kernel is higher than that with linear kernel. Similar results were obtained for other datasets when comparing Gaussian kernels with linear kernels. Therefore, we chose Gaussian kernel to train RT-SVR model for each chromatographic run.

The determination of hyper-parameters is complicated when training a SVR model. Caution is necessary to choose appropriate hyper-parameters in order to establish a good-performance model. 10-fold cross validation (CV) strategy was used to help choose the values of hyper-parameters. The optimal complexity parameter c was chosen from the range of 1.0 to 10.0 for linear kernel and from 1.0 to 100.0 for Gaussian kernel. The filter types used for both kernels are standardized training data with a function of $\chi_{std}=\frac{\chi-\chi_{\min}}{\chi_{\max}-\chi}$. The exponent p associated with kernel function is set to 1.0 while the γ in Gaussian model is set to 0.01 because they perform best in all cases by comparing to other values. We used all default values for other hyper-parameters since they did not yield significantly different results.

Some models have been published to study the property of retention time such as linear regressor by Krokhin [14], ANN by Petritis [16] etc. Linear regressor relates the retention time as the function of hydrophobicity. Although it is simple and ready to be used for small dataset, it is restricted to specific chromatographic conditions. Moreover, when applied to peptide identification, the method to choose an arbitrary RT threshold may cause either more false positives or identification loss. ANN method which employed a very large size of dataset (345000 unique peptides) for training is impractical when applied to protein profiling study. As a comparison, we ran Krokhin's linear model (SSRC)

(http://hs2.proteome.ca/SSRCalc/SSRCalc32.html, version 3.2) on our training data and tested the performance. The training dataset was used to calculate relative hydrophobicity with arbitrary value of 1.0 assigned to the two parameters a and b for each peptide in the dataset, followed by a plot of experimental retention time versus relative hydrophobicity. In this manner, the two parameters a and b were trained and used for test datasets to obtain the predicted retention times for the peptides in test set. Finally, the correlation coefficient R was calculated from the predicted and experimental retention times. Given that our chromatographic condition (300 Å with formic acid) is different from the four provided ones, we tested two conditions ("300Å column with TFA conditions" and "formic acid") and found that R values are larger with "formic acid" condition. Therefore, we set these larger R values as the test results from the SSRC model. The results for comparison are summarized in Table 2, from which we can see that in most cases, both Gaussian and linear RT-SVR outperform SSRC and only for run #10 the SSRC has a comparable performance to RT-SVR. This observation suggests that the RT-SVR model prefers relatively large dataset while the SSRC model works better on small dataset. In addition, it is worth noting that the RT-SVR model performance is not proportional to the size of training dataset. It only depends upon the diversity of dataset. For instance, for Gaussian kernel RT-SVR, R value (0.926) from dataset #6 (323 training examples) is smaller than that (0.956) from dataset #4 (171 training examples). This indicates that the data in dataset #6 are more diverse. This property is consistent with Klammer's observation [22].

We also studied if the RT-SVR model is robust with the size of training dataset. A graph of R versus the training dataset size was generated based on all 10 datasets, shown in Figure 3. RT-SVR maintains a stable performance with a standard deviation of 0.032 for Gaussian kernel and 0.053 for linear kernel, while R value for SSRC changes dramatically with a standard deviation of 0.099. Based on these observations, a conclusion can be made that the RT-SVR exhibits a stable performance if the size of training dataset is large enough for model construction.

3.2. Factors that affect the RT-SVR model performance

When we train the RT-SVR model, several concerns need to be addressed that could affect the performance of the model. The first concern is whether we can use multiple PSMs rather than unique peptides as examples in training and test datasets. Previous study [22] suggested that it is necessary to eliminate the redundancy of examples and only allow unique peptide sequence to be present in a dataset, which can avoid bias in the regression. However, given the nature of retention property of a compound on an LC column, we believe that there is a benefit to consider redundant PSMs. Theoretically, the elution peak of a given compound follows Gaussian distribution and spans duration from seconds to minutes when eluting with liquid chromatography. Accordingly, when the downstream mass spectrometer samples the compound and produces the final TIC (Total Ion Chromatogram), multiple mass spectra corresponding to the same compound will be acquired. Therefore, all retention times associated with these mass spectra represent the same peptide and should be all included for the RT-SVR model construction. Inspired by this chemical property, we choose multiple PSMs rather than unique peptide to produce training and test datasets. It is worth noting that this training technique will not lead to overfitting because the 45-element vectors corresponding to these PSMs are not the same due to different retention times. We took dataset #1 to compare the performance of RT-SVR models trained with multiple PSMs or unique peptides. The RT-SVR model trained with our method exhibits a higher R value of 0.964 compared to 0.792 with unique peptide training set (both are with Gaussian kernel). Using multiple PSMs to train a RT-SVR model represents a closer approximation to chemical condition of LC-MS based proteomic analysis. Thus, the resulting model has a superior performance to that trained with unique peptides.

Another concern is whether the correct rate of protein identifications in the training data significantly affects the performance of the RT-SVR model. Previous studies manually checked the original dataset to extract confident peptide identifications. By setting a high value of significance level, i.e., 10% FDR, a large source dataset was obtained to produce enough training examples [22, 23]. However, this method may result in less confident training examples which lead to a bias in regression. One way to address this issue is to use more confident source data to produce training and test datasets. This can be done by using Mascot results obtained at a FDR of 1% to form training and test datasets as we adopted in this study. To prove this, we compiled a set of training and test datasets which were randomly extracted from the Mascot results obtained at a FDR of 9.6% (run #5), followed by a RT-SVR with Gaussian kernel being trained and tested. We repeated 5 times and used the average R value to assess the model performance. Consequently, the RT-SVR model trained with the large dataset (430 examples) from less confident Mascot results (FDR=9.6%) produces R value of 0.864, compared to 0.959 produced by its counterpart (trained with 200 examples at FDR=1%). This comparison indicates that accuracy of training data is more important to ensure the good performance of a RT-SVR model.

Another major factor to consider is dataset size because the sample space and data diversity shows significant effect on model performance [30]. To investigate the relationship between the size of training dataset and model performance, we compiled a series of training datasets and a fixed test dataset which are randomly extracted from the dataset #1 obtained at a FDR of 1%. All datasets only contain confident peptide identifications with a Mascot score above Mascot Identity Threshold. A fixed dataset consisting of 101 peptides was first extracted and used to test the model performance. The training datasets with different sizes were then randomly extracted from the remaining part of the source dataset #1. The sizes of training datasets were designed as 50, 100, 150, 200 and 250. Three replicates of each training dataset were produced to eliminate random error due to data variation. A RT-SVR model was constructed by every specific training dataset and then tested with the fixed test dataset. The resulting R values were grouped by the size of training dataset and averaged arithmetically. Figure 4 shows the average R values with the associated standard deviations as error bar versus dataset size. As shown the R values corresponding to 50 and 100 training sets are 0.640 and 0.822, respectively, whereas all others are larger than 0.900. According to this observation, we can say that a size of 150 for training set is large enough to produce a satisfactory model performance ($R=0.901\pm0.013$). To strengthen this point, we investigated the quality of protein predictions. The trained RT-SVR models in previous step were applied to the application dataset, here the application dataset run #1 (p=0.10 or FDR=7.5%), to predict proteins with a q value of 0.01. Mascot (MIH) identified 135 proteins from this dataset at a FDR of 1%, which will be used as a comparison. The results are also shown in Figure 4, from which we can observe the trend of protein identifications over the size of training dataset. The model trained with the dataset containing 150 examples predicted 133 proteins, which is comparable to Mascot results. By considering both R value and protein prediction, we can conclude that a training dataset containing 150 or more peptides is good enough for RT-SVR model construction and the subsequent protein identification. Our study conducted on chromatographic run #9 and #10 proved this point. The RT-SVR model trained with a dataset consisting of 153 examples (dataset #9) yielded an R value of 0.903. This model enabled prediction of 108 proteins when being applied to dataset #9, in contrast with 94 identified by Mascot. However, with the RT-SVR model (R value is 0.874) trained with 57 peptides in chromatographic #10, only 40 proteins were identified in comparison to 44 proteins discovered by Mascot.

3.3 The contribution from features to retention time

According to previous studies[14], retention time of a peptide is determined by hydrophobicity of each amino acid residue, retaining property at C- or N-termini, structure and size of the peptide. Given the complex factors that could impact on retention time, it is meaningful to evaluate the overall contribution of each feature we used in this study to allow fine tune of the prediction model. Although we did not employ linear-kernel RT-SVR for protein identification, this model provides a benefit for the study of the contribution of every feature to retention time. It is worth noting that "retention contribution" here is different from the traditionally defined "hydrophobicity". Figure 5 demonstrates the contributions of several features including 20 amino acid residues, C-terminal R or K, length and mass of peptides. N-terminal amino acid residues were not shown because they exhibited similar contributions to the individual residues. Thus, the retention properties of N-terminal residues can be represented by those individual amino acid residues. The retention contributions were extracted from linear-kernel RT-SVR trained with dataset run #1 to #9, respectively. As shown in top panel, the average retention contributions with associated standard deviations as error bars are plotted with the features. Hydrophobic residues such as I, L, F and W have greater positive contributions while hydrophilic ones like K and R provides greater negative contributions. Although there is minimal contribution from C-terminal R or K, the length and mass make greater positive contributions. These observations are consistent with the theoretical prediction: large, more hydrophobic molecules tend to be retained longer as separated with RPLC. The bottom panel shows individual retention contributions from features for different chromatographic runs. As expected, although similar trends have been observed, changes are not consistent across varying chromatographic runs. This observation suggests that it is necessary to train a specific RT-SVR model for every chromatographic run.

3.4. Application of RT-SVR to protein identification

For massive proteomic data analysis, Mascot is commonly employed as one of the most popular database search engines for peptide and protein identifications. Mascot scores the similarity between the experimental MS/MS spectrum of a precursor and the theoretical MS/MS spectrum of a database sequence and reports peptide-spectrum matches with the highest score. Mascot identity threshold (MIT) or homology threshold (MHT) is used as the cut-off score threshold to filter out low-confidence peptide identifications. The peptide identifications are reported as confident target or decoy PSMs if their scores exceed a threshold, and then FDR is estimated as the fraction of the number of false positives (confident decoy PSMs) in the total number of positives (confident target PSMs). A customer-defined parameter, called significance level p, can be adjusted to obtain different FDRs and positive peptide identifications as well. In proteomic studies, a FDR around 1% is generally accepted as confident identifications. Although this method of FDR estimation is easy to follow, it may encounter a counterintuitive problem: an identical FDR may lead to different number of PSMs[27]. This results in unreliable assessment of peptide identifications using conventional Mascot search strategy. In addition, algorithmic bias on scoring PSMs may also lead to wrong peptide predictions.

RT-SVR uses retention time as an additional useful feature to reevaluate these peptide identifications. It will not only re-rank high-confidence PSMs but also rescue low-confidence PSMs. Instead of using FDR, we adopted a specific q value metric which is facilitated with retention time deviation to assess statistical significance of peptide identifications. Although q value estimation is widely used to assess the database search results, it has never been used in the retention time based machine learning technique such as the RT-SVR model described in this study. To the best of our knowledge, our work

represents the first attempt to introduce and implement the concept of q value estimation to this unique application.

For proteomic studies the RT-SVR model needs to be implemented for the application dataset (Mascot results) produced from the same chromatographic run that is used for model construction. To increase data space and accommodate more low-confidence PSMs to be rescued, the application dataset needs to be obtained from Mascot results at a high FDR (e.g. FDR ~10%). In this study, we ran a simultaneous target-decoy Mascot database search on all chromatographic runs (database search criteria are the same except precursor tolerance is 1.2 Da) with a p value of 0.10 (approximately FDR ~10%) to produce all application datasets (from #1 to #10). We then applied every RT-SVR model to the corresponding dataset from #1 to #9 (dataset #10 is used as performance comparison only) and generated the retention time deviation for each PSM in each dataset. q value was then calculated for each target PSM followed by screening of PSMs at a q value threshold of 0.01. The results are summarized in Table 3. As expected, the number of PSMs and proteins identified by RT-SVR are larger than those by Mascot (MIT) for each dataset, which is due to the fact that some low-confidence PSMs have been recovered by RT-SVR. Figure 6 shows the comparison of the number of confident PSMs obtained by RT-SVR method and Mascot identity threshold (MIT), respectively. As seen, the number of predicted PSMs increases by a range of values from 18.7% (as in dataset #6) to 165% (as in dataset #8), which corresponds to an increase of 4.3% and 89.4% additional unique peptides, respectively.

The use of q value metric other than FDR resolves the counterintuitive issue. Figure 7 shows the comparison of q value and FDR for the statistical validation of the same dataset #2. Dotted lines show the PSMs obtained from MIT and MHT over FDR while the solid line represents the PSMs over q value. A counterintuitive issue can be seen around ~1% FDR at which two different numbers of PSMs correspond to the same FDR value. As a comparison, no ambiguity issue is found for q value assessment. This is because the ambiguity is resolved mathematically by q value that is defined as the minimal FDR for any specific PSM. The primary distinction between the FDR and the q value is that the former is a property of a set of PSMs, whereas the latter is a property of a single PSM. We can therefore associate a unique q value with every target PSM in our dataset. It is worth noting that q value is equivalent to FDR because it is a statistical estimate based on the whole dataset although it is assigned to a single PSM [28, 33].

q value threshold is user-defined and can be adjusted (as p value in Mascot search) to result in different Δ RT thresholds. The distribution of q value over Δ RT for dataset #2 is plotted in Figure 8. As we can see, the specified q value of 0.01 defines a Δ RT threshold of ± 0.96 min, which identifies 419 target PSMs as the confident peptide identifications. If we increase the q value to 0.02, which defines a Δ RT threshold of ±4.8min, 978 PSMs would be identified. This flexibility of user-defined q value is beneficial when one wants to study more peptides with slightly higher false positive rate.

The use of q value as the cut-off threshold to assess the confidence of peptide identifications overcomes the limitations of using a fixed Δ RT threshold as in previous studies [22, 24]. Randomly choosing or training a fixed Δ RT threshold to cut off predicted peptides often results in more false positives or lose true positives. Here, the specified q value such as 0.01 defines the selection of a dynamic Δ RT threshold for a given dataset. Figure 9 demonstrates the selection of Δ RT threshold for all 9 application datasets. Datasets #3 and #6 show small Δ RT thresholds (±0.25 min and ±0.21min respectively), which correspond to a small number of PSMs identified. In contrast, datasets #1 and #8 have large Δ RT thresholds (± 2.91 min and ±2.53 min respectively), which lead to an increase of the number of peptide identifications. The different selection of Δ RT thresholds is determined by the performance

of the corresponding RT-SVR model and the quality of application dataset. Regardless of the Δ RT threshold, the accuracy of the peptide predictions is only determined by q value.

3.5 Inclusion of mascot identity threshold (MIT) into RT-SVR to increase prediction coverage

Table 3 shows that approximately 10% peptide identifications obtained by MIT method (at FDR 1%) fail to be identified by the RT-SVR model. This is in part due to model bias that considers some valid PSMs as outliers. To address this issue, we incorporated MIT into RT-SVR to obtain maximal coverage of peptide identifications. As two parallel approaches, MIT identifies peptides at a FDR of 1% while RT-SVR identifies peptides at a q value of 0.01 following the proposed protocols. Finally, we compile MIT and RT-SVR results together to produce the final report. By this means, we rescued all missed peptide identifications. Sample results from dataset #1 are provided in supplementary Table S3.

3.6 General applicability of the RT-SVR to proteomic data acquired on other MS platforms

In order to investigate the general applicability of our approach to shotgun proteomics data obtained using other MS instruments, we selected datasets acquired on a Thermo LTQ mass spectrometer due to its widespread use in large-scale proteomics studies. Two duplicate datasets of Yeast shotgun proteomics raw data were downloaded from PeptideAtlas (PAe001337). The details about raw data, search parameters and Trans-Proteomic Pipeline (TPP) results can be seen via ftp://ftp.peptideatlas.org/pub/PeptideAtlas/Repository/PAe001337. The raw data were converted into mgf data with MM file conversion tool (http://www.massmatrix.net/mm-cgi/downloads.py). The resulting mgf data were then searched with Mascot against SwissProt yeast database. We used the same search parameters as those shown online. Following the protocols we proposed in this study, for each dataset, we constructed a RT-SVR model based on Mascot results at FDR~1% and then applied the model to the larger Mascot searching results at p=0.10. In total, the RT-SVR resulted in 566 unique protein identifications from two duplicate datasets at a q value of 0.01, while Mascot (MIH) predicted 470 unique proteins at a FDR of 1% and TPP reported 499 unique proteins with a probability filter of 0.010. The proteins identified by RT-SVR are summarized in supplementary Table S4. Thus, we believe that RT-SVR method can also offer superior performance for MS-based proteomic analysis using LTQ platform.

4. Conclusions

Retention time, as one of the characteristic properties of a peptide sequence, can be used for distinguishing the confident peptide identifications from the false positives. Models that are built to predict retention time often fail to perform as expected if one simply assumes a linear correlation between RT and other intrinsic properties of a peptide. Support vector regressor with a Gaussian kernel is superior to others due to nonlinear regression and small dataset requirements. The peptides in training and test datasets affect the performance of a SVR model. Correct peptide identifications extracted from Mascot results at a FDR of 1% are favored. Given that Gaussian distribution regulates peptide separation and MS acquisition, multiple PSMs rather than unique peptides are chosen to produce training dataset and a better RT-SVR model can be trained. The size of training dataset also affects the performance of the RT-SVR model. According to our study, a dataset consisting of 150 confident peptides will lead to a high performance model ($R > 0.900$). Given the variation of retention time contributions from the features with chromatographic conditions, a specific RT-SVR model is needed for each chromatographic run.

When applying the trained RT-SVR model to a real dataset for peptide or protein identification, the main issue is how to accurately and effectively assess the confidence of the predictions. The method of choice in this study is q value estimation. According to our knowledge, this is the first report on the application of q value assessment to RT prediction. q value is calculated based on target-decoy search and resolves the counterintuitive issue caused by traditional FDR method. With q value assessment, peptides can be unambiguously identified. In this study, by using q value assessment, we have shown that RT-SVR model substantially outperforms Mascot, in the best case identifying 89.4% more peptides and 83.5% more proteins than using Mascot identity threshold at a q value of 0.01. These results suggest that the RT-SVR model with q value assessment has a great potential as a post-Mascot analysis tool to improve protein identifications in shotgun proteomics. The software package is written in Java and the windows-based graphical user interface can be freely downloaded from http://pages.cs.wisc.edu/~yadi/bioinfo/rtsvr/rtsvr.html.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported in part by a Department of Defense Pilot Award (W81XWH-11-1-0181), National Science Foundation (CHE-0957784), and National Institutes of Health through grant 1R01DK071801. L.L. acknowledges a Vilas Associate Award and an H.I. Romnes Research Fellowship.

ABBREVIATIONS

NKL	natural killer leukemia cell line
RPLC	reversed phase liquid chromatography
MS/MS	tandem mass spectrometry
PSM	peptide spectrum match
RT	retention time
Δ RT	deviation between predicted and experimental RT
SVR	support vector regressor
FDR	false discovery rate
MIT	Mascot identity threshold
MHT	Mascot homology threshold

References

1. Hunt DF, Michel H, Dickinson TA, Shabanowitz J, Cox AL, Sakaguchi K, et al. Peptides presented to the immune system by the murine class II major histocompatibility complex molecule I-Ad. Science. 1992; 256:1817–20. [PubMed: 1319610]
2. Wolters DA, Washburn MP, Yates JR 3rd. An automated multidimensional protein identification technology for shotgun proteomics. Anal Chem. 2001; 73:5683–90. [PubMed: 11774908]
3. Foster LJ, de Hoog CL, Zhang Y, Xie X, Mootha VK, Mann M. A mammalian organelle map by protein correlation profiling. Cell. 2006; 125:187–99. [PubMed: 16615899]
4. Eng JK, Mccormack AL, Yates JR. An Approach to Correlate Tandem Mass-Spectral Data of Peptides with Amino-Acid-Sequences in a Protein Database. J Am Soc Mass Spectr. 1994; 5:976–89.

5. Perkins DN, Pappin DJ, Creasy DM, Cottrell JS. Probability-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis. 1999; 20:3551–67. [PubMed: 10612281]
6. Craig R, Beavis RC. TANDEM: matching proteins with tandem mass spectra. Bioinformatics. 2004; 20:1466–7. [PubMed: 14976030]
7. Geer LY, Markey SP, Kowalak JA, Wagner L, Xu M, Maynard DM, et al. Open mass spectrometry search algorithm. J Proteome Res. 2004; 3:958–64. [PubMed: 15473683]
8. Elias JE, Haas W, Faherty BK, Gygi SP. Comparative evaluation of mass spectrometry platforms used in large-scale proteomics investigations. Nat Methods. 2005; 2:667–75. [PubMed: 16118637]
9. Cargile BJ, Bundy JL, Stephenson JL Jr. Potential for false positive identifications from large databases through tandem mass spectrometry. J Proteome Res. 2004; 3:1082–5. [PubMed: 15473699]
10. Qian WJ, Liu T, Monroe ME, Strittmatter EF, Jacobs JM, Kangas LJ, et al. Probability-based evaluation of peptide and protein identifications from tandem mass spectrometry and SEQUEST analysis: the human proteome. J Proteome Res. 2005; 4:53–62. [PubMed: 15707357]
11. Petritis K, Kangas LJ, Yan B, Monroe ME, Strittmatter EF, Qian WJ, et al. Improved peptide elution time prediction for reversed-phase liquid chromatography-MS by incorporating peptide sequence information. Anal Chem. 2006; 78:5026–39. [PubMed: 16841926]
12. Mant CT, Zhou NE, Hodges RS. Correlation of protein retention times in reversed-phase chromatography with polypeptide chain length and hydrophobicity. J Chromatogr. 1989; 476:363–75. [PubMed: 2777984]
13. Mant CT, Hodges RS. Context-dependent effects on the hydrophilicity/hydrophobicity of side-chains during reversed-phase high-performance liquid chromatography: Implications for prediction of peptide retention behaviour. J Chromatogr A. 2006; 1125:211–9. [PubMed: 16814308]
14. Krokhin OV, Craig R, Spicer V, Ens W, Standing KG, Beavis RC, et al. An improved model for prediction of retention times of tryptic peptides in ion pair reversed-phase HPLC: its application to protein peptide mapping by off-line HPLC-MALDI MS. Mol Cell Proteomics. 2004; 3:908–19. [PubMed: 15238601]
15. Baczek T, Wiczling P, Marszall M, Heyden YV, Kaliszan R. Prediction of peptide retention at different HPLC conditions from multiple linear regression models. J Proteome Res. 2005; 4:555–63. [PubMed: 15822934]
16. Petritis K, Kangas LJ, Ferguson PL, Anderson GA, Pasa-Tolic L, Lipton MS, et al. Use of artificial neural networks for the accurate prediction of peptide liquid chromatography elution times in proteome analyses. Anal Chem. 2003; 75:1039–48. [PubMed: 12641221]
17. Shinoda K, Sugimoto M, Yachie N, Sugiyama N, Masuda T, Robert M, et al. Prediction of liquid chromatographic retention times of peptides generated by protease digestion of the Escherichia coli proteome using artificial neural networks. J Proteome Res. 2006; 5:3312–7. [PubMed: 17137332]
18. Palmblad M, Ramstrom M, Markides KE, Hakansson P, Bergquist J. Prediction of chromatographic retention and protein identification in liquid chromatography/mass spectrometry. Anal Chem. 2002; 74:5826–30. [PubMed: 12463368]
19. Kawakami T, Tateishi K, Yamano Y, Ishikawa T, Kuroki K, Nishimura T. Protein identification from product ion spectra of peptides validated by correlation between measured and predicted elution times in liquid chromatography/mass spectrometry. Proteomics. 2005; 5:856–64. [PubMed: 15668996]
20. Palmblad M, Ramstrom M, Bailey CG, McCutchen-Maloney SL, Bergquist J, Zeller LC. Protein identification by liquid chromatography-mass spectrometry using retention time prediction. J Chromatogr B Analyt Technol Biomed Life Sci. 2004; 803:131–5.
21. Strittmatter EF, Kangas LJ, Petritis K, Mottaz HM, Anderson GA, Shen Y, et al. Application of peptide LC retention time information in a discriminant function for peptide identification by tandem mass spectrometry. J Proteome Res. 2004; 3:760–9. [PubMed: 15359729]

22. Klammer AA, Yi X, MacCoss MJ, Noble WS. Improving tandem mass spectrum identification using peptide retention time prediction across diverse chromatography conditions. Anal Chem. 2007; 79:6111–8. [PubMed: 17622186]

23. Le Bihan T, Robinson MD, Stewart, Figeys D. Definition and characterization of a “trypsinosome” from specific peptide characteristics by nano-HPLC-MS/MS and in silico analysis of complex protein mixtures. J Proteome Res. 2004; 3:1138–48. [PubMed: 15595722]

24. Krokhin OV. Sequence-specific retention calculator. Algorithm for peptide retention prediction in ion-pair RP-HPLC: application to 300- and 100-A pore size C18 sorbents. Anal Chem. 2006; 78:7785–95. [PubMed: 17105172]

25. Storey JD, Xiao W, Leek JT, Tompkins RG, Davis RW. Significance analysis of time course microarray experiments. Proc Natl Acad Sci U S A. 2005; 102:12837–42. [PubMed: 16141318]

26. Kall L, Storey JD, MacCoss MJ, Noble WS. Posterior error probabilities and false discovery rates: two sides of the same coin. J Proteome Res. 2008; 7:40–4. [PubMed: 18052118]

27. Kall L, Storey JD, MacCoss MJ, Noble WS. Assigning significance to peptides identified by tandem mass spectrometry using decoy databases. J Proteome Res. 2008; 7:29–34. [PubMed: 18067246]

28. Kall L, Canterbury JD, Weston J, Noble WS, MacCoss MJ. Semi-supervised learning for peptide identification from shotgun proteomics datasets. Nat Methods. 2007; 4:923–5. [PubMed: 17952086]

29. Brosch M, Yu L, Hubbard T, Choudhary J. Accurate and sensitive peptide identification with Mascot Percolator. J Proteome Res. 2009; 8:3176–81. [PubMed: 19338334]

30. Moruz L, Tomazela D, Kall L. Training, selection, and robust calibration of retention time models for targeted proteomics. J Proteome Res. 2010; 9:5209–16. [PubMed: 20735070]

31. Huttlin EL, Hegeman AD, Harms AC, Sussman MR. Prediction of error associated with false-positive rate determination for peptide identification in large-scale proteomics experiments using a combined reverse and forward peptide sequence database strategy. J Proteome Res. 2007; 6:392–8. [PubMed: 17203984]

32. Bland JM, Altman DJ. Regression analysis. Lancet. 1986; 1:908–9. [PubMed: 2870372]

33. Spivak M, Weston J, Bottou L, Kall L, Noble WS. Improvements to the percolator algorithm for Peptide identification from shotgun proteomics data sets. J Proteome Res. 2009; 8:3737–45. [PubMed: 19385687]

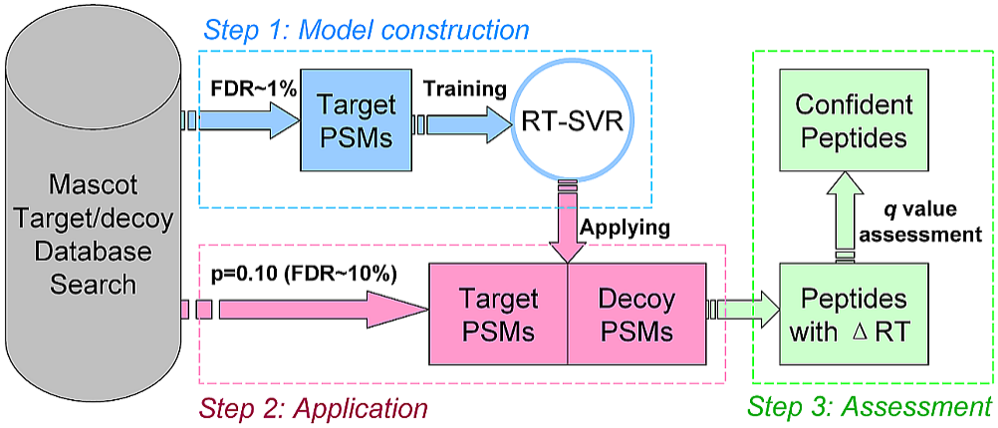


Figure 1.

The workflow of RT-SVR in processing proteomic data. The first step is to construct the RT-SVR model. Target PSMs (at 1%FDR) are screened (remove those with Mascot score < Mascot identify threshold) to create training and testing datasets (split with a ratio of 3:1). The second step is to apply the trained RT-SVR to Mascot results (at p=0.10) so as to filter out false positive predictions. Both target and decoy PSMs are processed with RT-SVR following by q value assessment by which confident peptide predictions are selected out at a given q value.

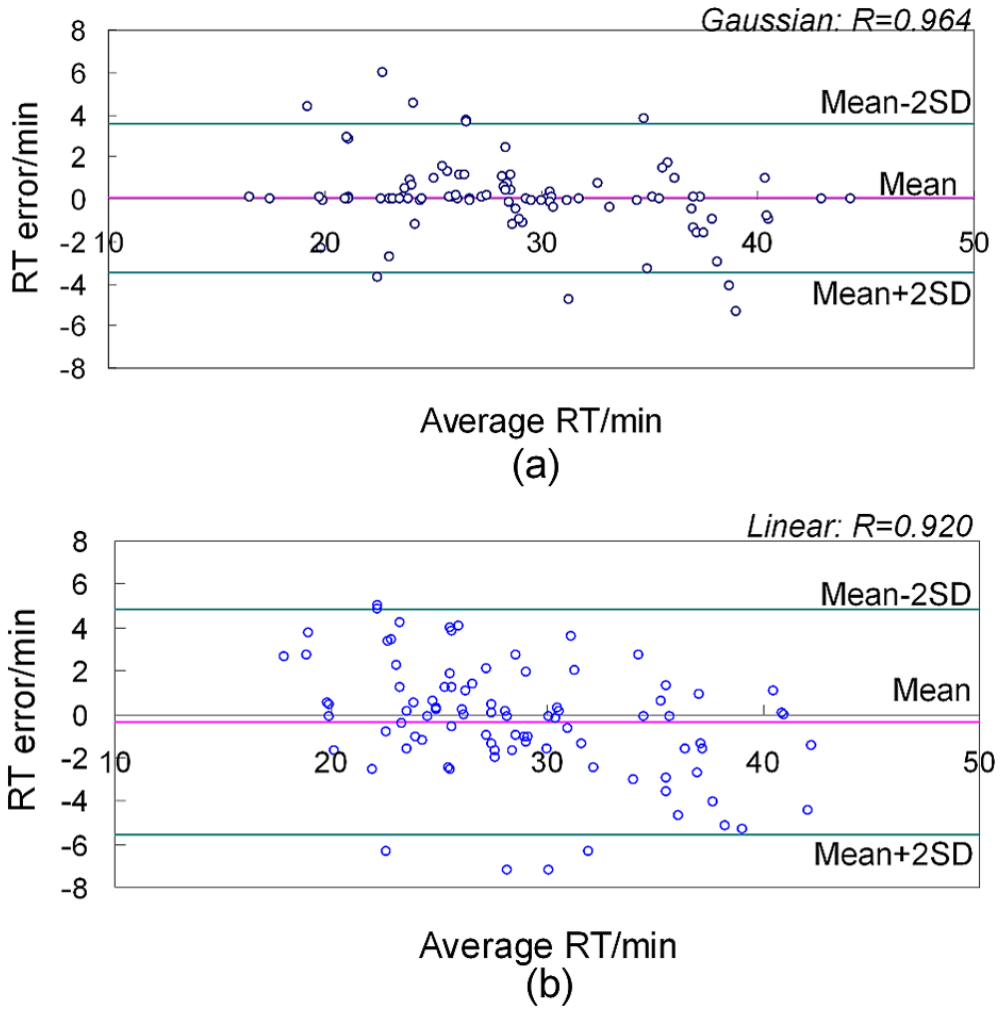


Figure 2.
The Bland-Altman plot shows the distribution of RT deviation over the average of the predicted and experimental RTs. The Pearson's correlation R value is specified at the top right. Dataset #1 was used to make graphs. a) Gaussian kernel; b) Linear kernel.

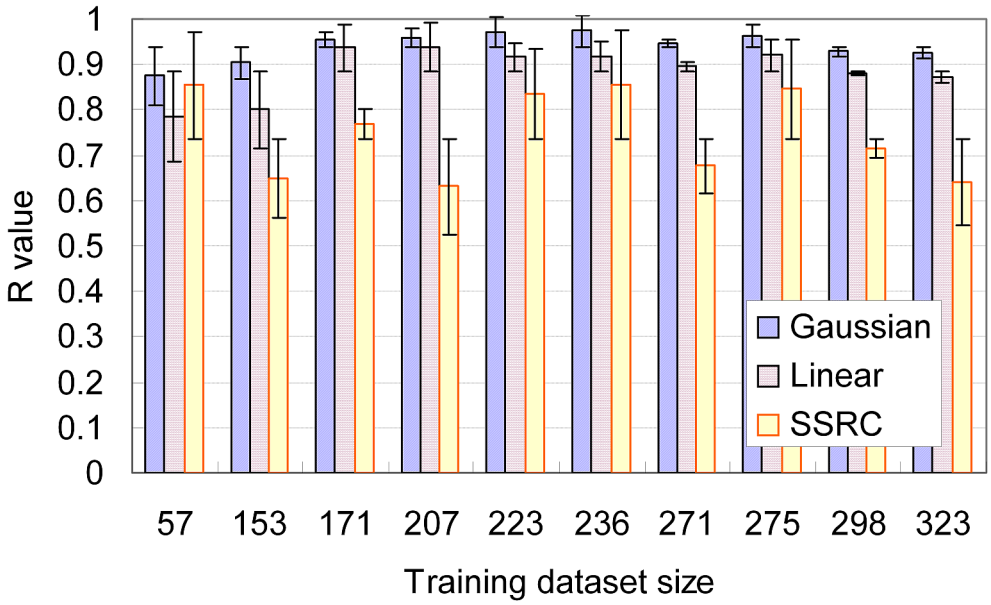


Figure 3.
The performance and robustness comparison for different RT-SVR models and SSRC linear regressors.

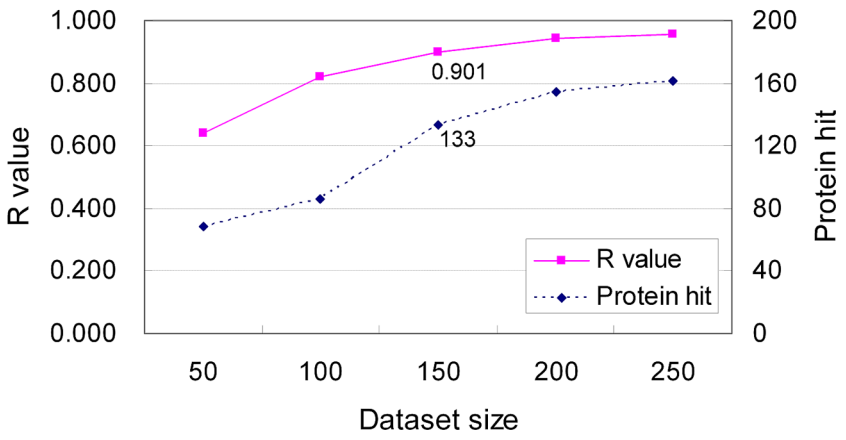


Figure 4.
The performance of RT-SVR model as a function of the number of training examples. Data are obtained from dataset #1. Each data point represents the average value of 3 replicates.

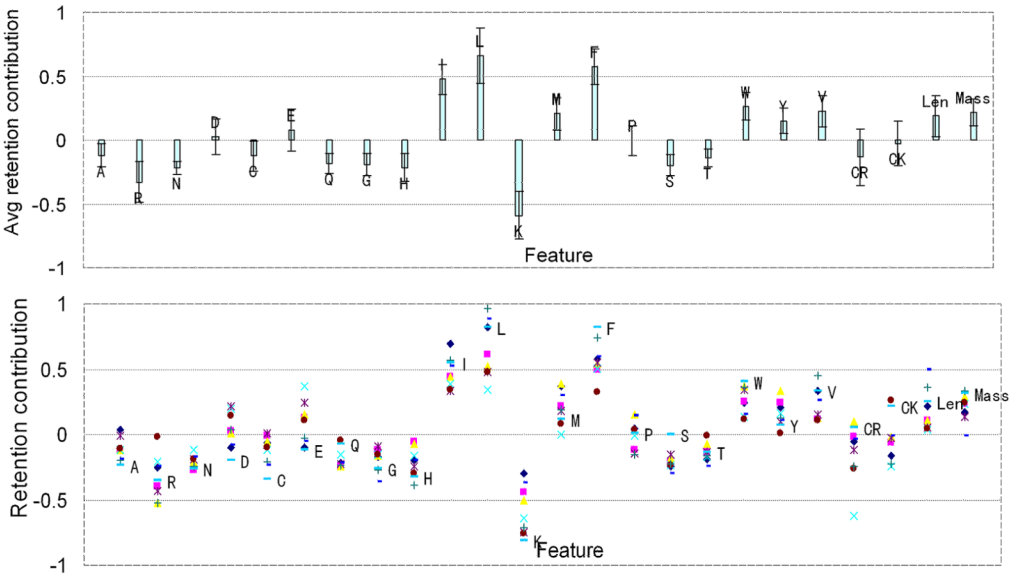


Figure 5.

Retention contributions of some features. Shown are the weights from linear-kernel RT-SVR corresponding to 20 amino acid residues, C-terminal R or K, length and mass. a) Average retention contribution (standard deviation shown as error bars); b) Individual contributions for each dataset.

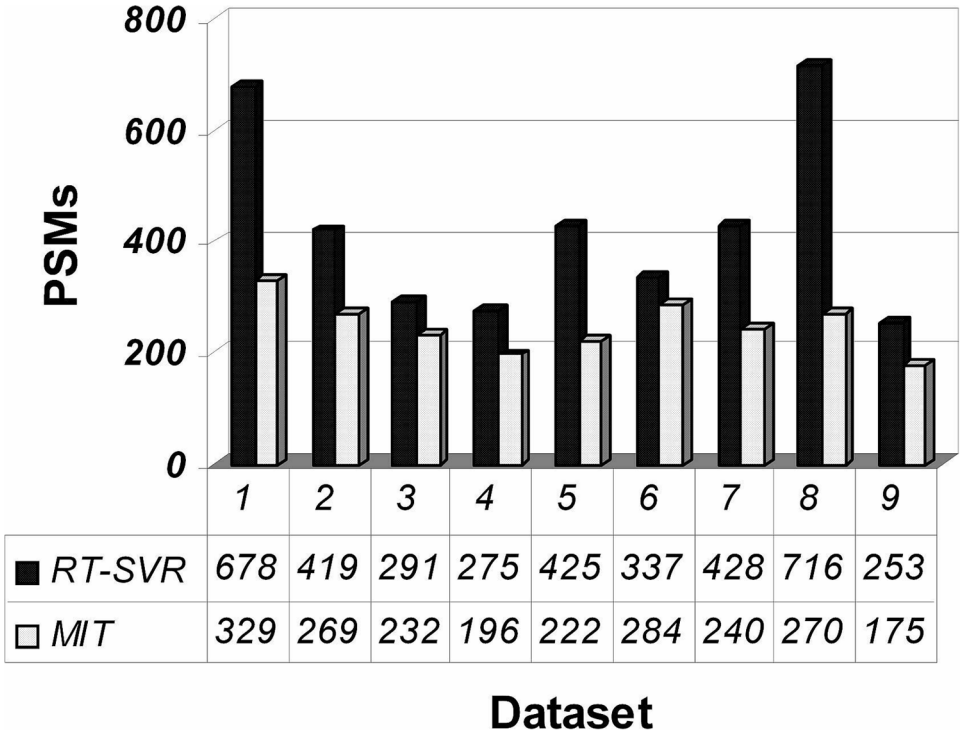


Figure 6.
The plots of PSMs versus q value or FDR. The curves of PSMs over FDR (for MIT and MHT) are counterintuitive around 1% FDR. The plot of PSMs over q value (for RT-SVR) resolves this issue. Comparison indicates that RT-SVR outperforms MIT and MHT. The results are obtained from application dataset #2.

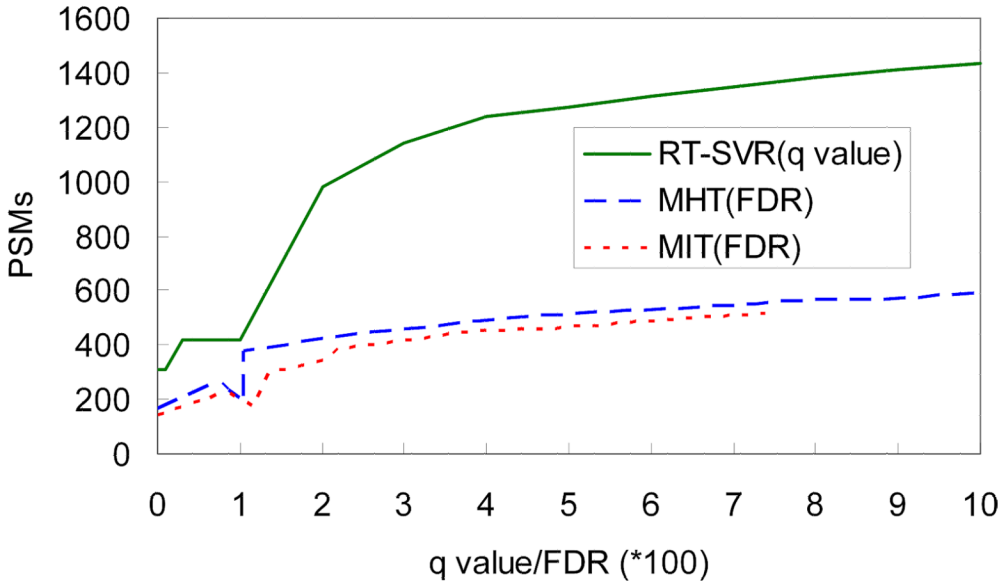


Figure 7.
The comparison of PSMs identified with RT-SVR and MIT at a q value of 0.01.

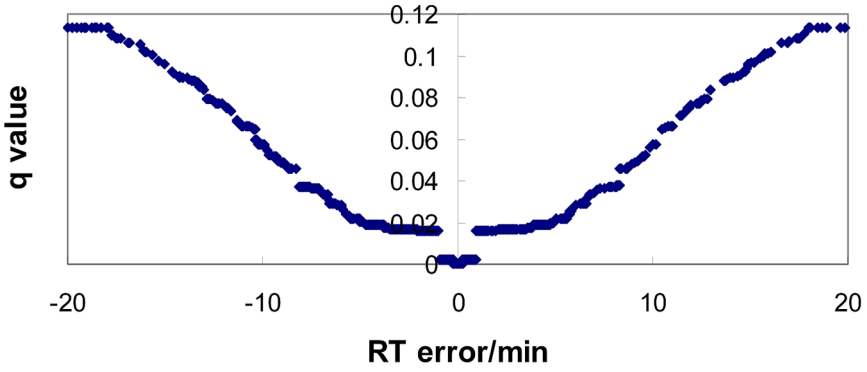


Figure 8.
The distribution of q value over RT error (Δ RT). Δ RT threshold depends on q value. Data obtained from dataset #2.

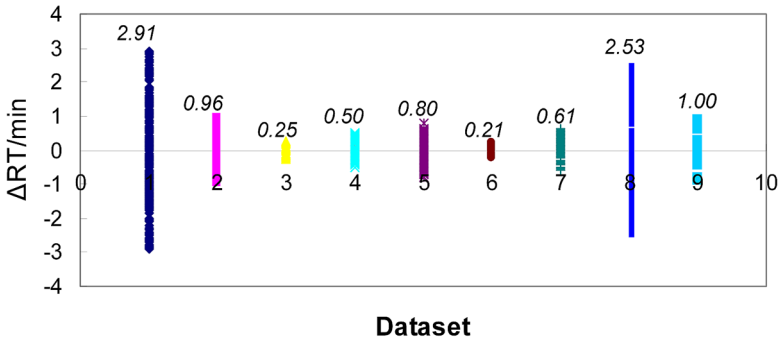


Figure 9.

The dynamic Δ RT thresholds for all 9 datasets are determined by a specified q value of 0.01. The number on top of each bar represents the upper bound of Δ RT threshold while the lower bound not shown has the same value but negative sign.

Table 1

The dataset used in this study. All data with scores exceeding identity threshold are extracted from Mascot results obtained at a FDR of 1%.

Run	Confident peptides	Training	Test
1	376	275	101
2	364	271	93
3	306	236	70
4	215	171	44
5	293	207	86
6	410	323	87
7	284	223	61
8	374	298	76
9	206	153	53
10	77	57	20

Table 2

The performance (R value) of RT-SVR and SSRC linear regressor for each dataset

Data set	RT-SVR		SSRC*
	Gaussian	Linear	
1	0.964	0.920	0.846
2	0.948	0.896	0.677
3	0.975	0.919	0.857
4	0.956	0.936	0.770
5	0.959	0.938	0.631
6	0.926	0.873	0.641
7	0.972	0.916	0.836
8	0.928	0.882	0.716
9	0.903	0.800	0.650
10	0.874	0.786	0.854

*SSRC is Sequence Specific Retention Calculator (http://hs2.proteome.ca/SSRCalc/SSRCalcHelp.htm, version 3.2)

Table 3

The results for peptide and protein identifications when applying RT-SVR model with q value assessment on each application dataset.

Run	PSM(FDR) p=0.10	Protein p=0.10	ΔRT/min	PSM (FDR=1%, q=0.01)			Protein (FDR=1%, q=0.01)		
				RT-SVR	MIT	Overlap%	RT-SVR	MIT	Overlap%
1	1488 (7.5%)	174	±2.91	678	329	95	158	135	96
2	1947 (7.4%)	195	±0.96	419	269	92	127	92	90
3	1508 (9.3%)	199	±0.25	291	232	87	124	103	89
4	1440 (9.2%)	172	±0.50	275	196	92	119	92	91
5	1906 (9.6%)	211	±0.80	425	222	80	147	129	86
6	2016 (9.5%)	191	±0.21	337	284	87	125	110	93
7	3430 (19.9%)	248	±0.61	428	240	93	152	120	93
8	2312 (10.9%)	209	±2.53	716	270	94	167	91	96
9	1261 (9.7%)	159	±1.00	253	175	87	108	94	89

The specified q value is 0.01. The application dataset is obtained from Mascot search at a p value of 0.10. The PSMs and protein hits from Mascot identity threshold (MIT) are at a FDR of 1%. The PSMs and protein hits from RT-SVR are at a q value of 0.01. Overlap% is calculated as the number of identifications that are identified by both Mascot and RT-SVR divided by the total number of identifications by Mascot only.

glycosphingolipid-enriched microdomains of rat natural killer cells. Proteomics 2005;5:113–22.
[24] Hanna J, Fitchett J, Rowe T, Daniels M, Heller M, Gonen-Gross T, et al. Proteomic analysis of human natural killer cells: insights on new potential NK immune functions. Mol Immunol 2005;42:425–31.
[25] Hanna J, Gonen-Gross T, Fitchett J, Rowe T, Daniels M, Arnon TI, et al. Novel APC-like properties of human NK cells directly regulate T cell activation. J Clin Invest 2004;114:1612–23.
[26] Casey TM, Meade JL, Hewitt EW. Organelle proteomics: identification of the exocytic machinery associated with the natural killer cell secretory lysosome. Mol Cell Proteomics 2007;6:767–80.
[27] Gubbels JA, Felder M, Horibata S, Belisle JA, Kapur A, Holden H, et al. MUC16 provides immune protection by inhibiting synapse formation between NK and ovarian tumor cells. Mol Cancer 2010;9:11.
[28] Belisle JA, Gubbels JA, Raphael CA, Migneault M, Rancourt C, Connor JP, et al. Peritoneal natural killer cells from epithelial ovarian cancer patients show an altered phenotype and bind to the tumour marker MUC16 (CA125). Immunology 2007;122: 418–29.
[29] Patankar MS, Jing Y, Morrison JC, Belisle JA, Lattanzio FA, Deng Y, et al. Potent suppression of natural killer cell response mediated by the ovarian tumor marker CA125. Gynecol Oncol 2005;99:704–13.
[30] Song C, Ye M, Han G, Jiang X, Wang F, Yu Z, et al. Reversed-phase-reversed-phase liquid chromatography approach with high orthogonality for multidimensional separation of phosphopeptides. Anal Chem 2010;82:53–6.
[31] Brosch M, Yu L, Hubbard T, Choudhary J. Accurate and sensitive peptide identification with Mascot Percolator. J Proteome Res 2009;8:3176–81.
[32] Zhang Y, Wen Z, Washburn MP, Florens L. Refinements to label free proteome quantitation: how to deal with peptides shared by multiple proteins. Anal Chem 2010;82:2272–81.
[33] Zybailov B, Mosley AL, Sardiu ME, Coleman MK, Florens L, Washburn MP. Statistical analysis of membrane proteome expression changes in Saccharomyces cerevisiae. J Proteome Res 2006;5:2339–47.
[34] Schmidt H, Gelhaus C, Nebendahl M, Lettau M, Watzl C, Kabelitz D, et al. 2-D DIGE analyses of enriched secretory lysosomes reveal heterogeneous profiles of functionally relevant proteins in leukemic and activated human NK cells. Proteomics 2008;8:2911–25.
[35] Gilar M, Olivova P, Daly AE, Gebler JC. Orthogonality of separation in two-dimensional liquid chromatography. Anal Chem 2005;77:6426–34.
[36] Strader MB, Tabb DL, Hervey WJ, Pan C, Hurst GB. Efficient and specific trypsin digestion of microgram to nanogram quantities of proteins in organic-aqueous solvent systems. Anal Chem 2006;78:125–34.
[37] Wang KS, Ritz J, Frank DA. IL-2 induces STAT4 activation in primary NK cells and NK cell lines, but not in T cells. J Immunol 1999;162:299–304.
[38] Wang KS, Frank DA, Ritz J. Interleukin-2 enhances the response of natural killer cells to interleukin-12 through up-regulation of the interleukin-12 receptor and STAT4. Blood 2000;95:3183–90.
[39] Koyama S. Augmented human-tumor-cytolytic activity of peripheral blood lymphocytes and cells from a mixed lymphocyte/tumor culture activated by interleukin-12 plus interleukin-2, and the phenotypic characterization of the cells in patients with advanced carcinoma. J Cancer Res Clin Oncol 1997;123:478–84.
[40] Gollob JA, Mier JW, Veenstra K, McDermott DF, Clancy D, Clancy M, et al. Phase I trial of twice-weekly intravenous interleukin 12 in patients with metastatic renal cell cancer or malignant melanoma: ability to maintain IFN-gamma induction is associated with clinical response. Clin Cancer Res 2000;6:1678–92.
[41] Andersen JN, Mortensen OH, Peters GH, Drake PG, Iversen LF, Olsen OH, et al. Structural and evolutionary relationships among protein tyrosine phosphatase domains. Mol Cell Biol 2001;21:7117–36.
[42] Simoncic PD, Lee-Loy A, Barber DL, Tremblay ML, McGlade CJ. The T cell protein tyrosine phosphatase is a negative regulator of Janus family kinases 1 and 3. Curr Biol 2002;12:446–53.
[43] Myers MP, Andersen JN, Cheng A, Tremblay ML, Horvath CM, Parisien JP, et al. TYK2 and JAK2 are substrates of protein-tyrosine phosphatase 1B. J Biol Chem 2001;276: 47771–4.
[44] Aoki N, Matsuda T. A cytosolic protein-tyrosine phosphatase PTP1B specifically dephosphorylates and deactivates prolactin-activated STAT5a and STAT5b. J Biol Chem 2000;275:39718–26.
[45] Kurki P, Lotz M, Ogata K, Tan EM. Proliferating cell nuclear antigen (PCNA)/cyclin in activated human T lymphocytes. J Immunol 1987;138:4114–20.
[46] Bohler T, Nolting J, Kamar N, Gurragchaa P, Reisener K, Glander P, et al. Validation of immunological biomarkers for the pharmacodynamic monitoring of immunosuppressive drugs in humans. Ther Drug Monit 2007;29:77–86.
[47] Barten MJ, Tarnok A, Garbade J, Bittner HB, Dhein S, Mohr FW, et al. Pharmacodynamics of T-cell function for monitoring immunosuppression. Cell Prolif 2007;40:50–63.
[48] Millan O, Benitez C, Guillen D, Lopez A, Rimola A, Sanchez-Fueyo A, et al. Biomarkers of immunoregulatory status in stable liver transplant recipients undergoing weaning of immunosuppressive therapy. Clin Immunol 2010;137:337–46.
[49] Niwa M, Miwa Y, Kuzuya T, Iwasaki K, Haneda M, Ueki T, et al. Stimulation index for PCNA mRNA in peripheral blood as immune function monitoring after renal transplantation. Transplantation 2009;87:1411–4.
[50] Rosental B, Brusilovsky M, Hadad U, Oz D, Appel MY, Afergan F, et al. Proliferating cell nuclear antigen is a novel inhibitory ligand for the natural cytotoxicity receptor NKp44. J Immunol 2011;187:5693–702.
[51] Rosental B, Hadad U, Brusilovsky M, Campbell KS, Porgador A. A novel mechanism for cancer cells to evade immune attack by NK cells: The interaction between NKp44 and proliferating cell nuclear antigen. Oncoimmunology 2012;1:572–4.
[52] Boles KS, Stepp SE, Bennett M, Kumar V, Mathew PA. 2B4 (CD244) and CS1: novel members of the CD2 subset of the immunoglobulin superfamily molecules expressed on natural killer cells and other leukocytes. Immunol Rev 2001;181:234–49.
[53] Vaidya SV, Stepp SE, McNerney ME, Lee JK, Bennett M, Lee KM, et al. Targeted disruption of the 2B4 gene in mice reveals an in vivo role of 2B4 (CD244) in the rejection of B16 melanoma cells. J Immunol 2005;174:800–7.
[54] Kim EO, Kim TJ, Kim N, Kim ST, Kumar V, Lee KM. Homotypic cell to cell cross-talk among human natural killer cells reveals differential and overlapping roles of 2B4 and CD2. J Biol Chem 2010;285:41755–64.
[55] Sinha SK, Gao N, Guo Y, Yuan D. Mechanism of induction of NK activation by 2B4 (CD244) via its cognate ligand. J Immunol 2010;185:5205–10.
[56] Gray JX, Haino M, Roth MJ, Maguire JE, Jensen PN, Yarme A, et al. CD97 is a processed, seven-transmembrane, heterodimeric receptor associated with inflammation. J Immunol 1996;157:5438–47.
[57] Kop EN, Matmati M, Pouwels W, Leclercq G, Tak PP, Hamann J. Differential expression of CD97 on human lymphocyte subsets and limited effect of CD97 antibodies on allogeneic T-cell stimulation. Immunol Lett 2009;123:160–8.
[58] Haynes BF, Hemler ME, Mann DL, Eisenbarth GS, Shelhamer J, Mostowski HS, et al. Characterization of a monoclonal

antibody (4 F2) that binds to human monocytes and to a subset of activated lymphocytes. J Immunol 1981;126: 1409–14.
[59] Moretta A, Mingari MC, Haynes BF, Sekaly RP, Moretta L, Fauci AS. Phenotypic characterization of human cytolytic T lymphocytes in mixed lymphocyte culture. J Exp Med 1981;153:213–8.
[60] Yan Y, Vasudevan S, Nguyen HT, Merlin D. Intestinal epithelial CD98: an oligomeric and multifunctional protein. Biochim Biophys Acta 2008;1780:1087–92.
[61] Deves R, Boyd CA. Surface antigen CD98(4 F2): not a single membrane protein, but a family of proteins with multiple functions. J Membr Biol 2000;173:165–77.
[62] Warren AP, Patel K, Miyamoto Y, Wygant JN, Woodside DG, McIntyre BW. Convergence between CD98 and integrin-mediated T-lymphocyte co-stimulation. Immunology 2000;99:62–8.
[63] Miyamoto YJ, Mitchell JS, McIntyre BW. Physical association and functional interaction between beta1 integrin and CD98 on human T lymphocytes. Mol Immunol 2003;39:739–51.
[64] Cantoni C, Bottino C, Augugliaro R, Morelli L, Marcenaro E, Castriconi R, et al. Molecular and functional characterization of IRp60, a member of the immunoglobulin superfamily that functions as an inhibitory receptor in human NK cells. Eur J Immunol 1999;29:3148–59.
[65] Ju X, Zenke M, Hart DN, Clark GJ. CD300a/c regulate type I interferon and TNF-alpha secretion by human plasmacytoid dendritic cells stimulated with TLR7 and TLR9 ligands. Blood 2008;112:1184–94.
[66] Clark GJ, Rao M, Ju X, Hart DN. Novel human CD4+ T lymphocyte subpopulations defined by CD300a/c molecule expression. J Leukoc Biol 2007;82:1126–35.
[67] Alvarez Y, Tang X, Coligan JE, Borrego F. The CD300a (IRp60) inhibitory receptor is rapidly up-regulated on human neutrophils in response to inflammatory stimuli and modulates CD32a (FcgammaRIIa) mediated signaling. Mol Immunol 2008;45:253–8.
[68] Nakahashi-Oda C, Tahara-Hanaoka S, Honda S, Shibuya K, Shibuya A. Identification of phosphatidylserine as a ligand for the CD300a immunoreceptor. Biochem Biophys Res Commun 2012;417:646–50.
[69] Deblandre GA, Marinx OP, Evans SS, Majjaj S, Leo O, Caput D, et al. Expression cloning of an interferon-inducible 17-kDa membrane protein implicated in the control of cell growth. J Biol Chem 1995;270:23860–6.
[70] Friedman RL, Manly SP, McMahon M, Kerr IM, Stark GR. Transcriptional and posttranscriptional regulation of interferon-induced gene expression in human cells. Cell 1984;38:745–55.
[71] Yang G, Xu Y, Chen X, Hu G. IFITM1 plays an essential role in the antiproliferative action of interferon-gamma. Oncogene 2007;26:594–603.
[72] Di Santo JP. Natural killer cell developmental pathways: a question of balance. Annu Rev Immunol 2006;24:257–86.